COMPUTATIONAL PREDICTION AND EXPERIMENTAL VALIDATION OF NOVEL MOUSE IMPRINTED GENES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Chelsea Marie McLean

August 2009

COMPUTATIONAL PREDICTION AND EXPERIMENTAL VALIDATION OF NOVEL MOUSE IMPRINTED GENES

Chelsea Marie McLean, Ph.D.

Cornell University 2009

Epigenetic modifications, including DNA methylation and covalent modifications to histone tails, are major contributors to the regulation of gene expression. These changes are reversible, yet can be stably inherited, and may last for multiple generations without change to the underlying DNA sequence. Genomic imprinting results in expression from one of the two parental alleles and is one example of epigenetic control of gene expression. So far, 60 to 100 imprinted genes have been identified in the human and mouse genomes, respectively. Identification of additional imprinted genes has become increasingly important with the realization that imprinting defects are associated with complex disorders ranging from obesity to diabetes and behavioral disorders. Despite the importance imprinted genes play in human health, few studies have undertaken genome-wide searches for new imprinted genes. These have used empirical approaches, with some success. However, computational prediction of novel imprinted genes has recently come to the forefront. I have developed generalized linear models using data on a variety of sequence and epigenetic features within a training set of known imprinted genes. The resulting models were used to predict novel imprinted genes in the mouse genome. After imposing a stringency threshold, I compiled an initial candidate list of 155 genes. A subset of these genes was

tested for evidence of imprinting using allele-specific restriction digests in either brain or placenta. Of the 10 genes tested in placenta, 2 showed evidence of maternal allele-specific expression. I also designed a custom microarray to test a total of 563 genes predicted as imprinted at lower stringency levels. Of these 563 genes, I experimentally tested 32 in placenta and 8 in brain, resulting in the identification of an additional 5 novel imprinted genes in placenta. This study is the first to demonstrate the utility of epigenetic marks in the prediction of imprinted genes. Furthermore, specific combinations of epigenetic marks were commonly found within particular regions relative to the transcriptional start sites of imprinted genes, implicating their placement and localization in the imprinting mechanism.

BIOGRAPHICAL SKETCH

Chelsea McLean was born on June 19, 1982 in Rochester, Michigan to Dara and Larry McLean. She graduated from Rochester Adams High School in 2000, where she took part in two field courses: Tropical Rainforest Ecology and Coral Reef Biology in Belize, Central America and The Ecology of Greater Yellowstone National Park in Yellowstone National Park. She credits participation in these courses for a large part of her interest in science. After high school, she continued on to the Lyman Briggs School at Michigan State University in East Lansing, Michigan. While at Michigan State, she participated in a wide variety of service organizations and taught two Honors level Biology labs: Lyman Briggs Honors Organismal Biology Lab, under the direction of Dr. Charles Elzinga, and Lyman Briggs Honors Cell and Molecular Biology Lab, under the direction of Dr. John Urbance. She also conducted research for two years in the Canine Genetics Lab of Dr. Patrick Venta and for one summer in the Small Fruit Entomology Lab of Dr. Rufus Isaacs. In 2004, Chelsea graduated with High Honor and a Bachelor of Science degree in Microbiology and Molecular Genetics and was recognized as a Michigan State University Outstanding Senior. After graduation, she moved to Ithaca, New York to attend graduate school at Cornell University and to pursue a Ph.D. in Genetics and Development under the instruction of Dr. Paul Soloway. Upon completion of her dissertation in 2009, Chelsea continued on to a postdoctoral position in the lab of Drs. Elizabeth Robertson and Elizabeth Bikoff at the University of Oxford in Oxford, U.K.

To my family, Dara, Larry, and Danielle: for your love and support. And for always telling me that I can do anything I set my mind to.

To my teachers: Todd Vince and Dave Glenn for sharing your enthusiasm for science and for being wonderful teachers.

To my best friend: Nick Brideau for always being there for me.

To all of my friends: For making my time in Ithaca memorable and enjoyable.

ACKNOWLEDGMENTS

There are so many people I have to acknowledge for their help, guidance, and friendship while I was at Cornell. Without these people, I would not be where I am today. First, my advisor, Paul Soloway, for his support over the last five years. From him, I learned countless scientific techniques, as well as how to think both critically and creatively. Also, the members of my special committee for their guidance throughout the many difficult phases of my various Ph.D. projects: Dr. Andrew Clark, Dr. John Schimenti, and Dr. Tudorita Tumbar.

And, along the way, I was lucky to have some wonderful labmates, who taught me so much through example and through discussion. Rebecca Holmes, Anders Lindroth, and Yoon Jung Park were an invaluable part of my experience at Cornell, as was Herry Herman, who was my mentor during my rotation in the Soloway lab. And, of course Patrick Murphy for his sunny disposition and his unbridaled enthusiasm for science, Ruquian Zhao for her support, encouragememt, and phoenix good luck charm. Jim Putnam for his ever-sunny attitude and for keeping our lab running smoothly.

And, of course, my friends. Karen Osorio for always listening, for making me laugh, and for her amazing desserts. Rebecca Holmes and Nathalie Nicod, who were there with me everyday and provided (decaf) coffee breaks, laughter, chocolate, and encouragement as needed. Sam and David Infanger for beautiful days on Ms. Green and fun nights of creative social events. Kim Holloway for her endless cheer, great barbecue, and love of good

fish and chips…with or without mushy peas. Amanda Larracuente for our fun dinners out and Bluestone Mojitos. And, of course, “The Girls” for our girls’ nights.

I also have to thank my very best friend and future husband, Nick Brideau. Without him, I could not have made it through everything that happened to me while I was in Ithaca, both personally and professionally. Without his love and support, I quite literally would not be writing this dissertation.

Finally, I would like to acknowledge my family. It is because of them that I learned the value of education and organization. From as far back as I can remember, my family managed to combine education and fun. For this I am grateful, as I never viewed education as a chore, but as a privelage. Also, because of the various after school activities that I was involved in, I learned how to multi-task and to complete projects under a deadline. I couldn’t have learned these skills without dedicated parents who not only shuttled me to and from my various activities on time, but also made sure my homework was complete at the end of each day. Even now, from across the country, my family manages to provide their unconditional support, for which I am extremely grateful, and I can only hope that I have made them proud.

TABLE OF CONTENTS

Appendix

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3'UTR – 3' untranslated region

5-azaC – 5-aza cytidine

5'UTR - 5' untranslated region

ADS – allele discriminatory signal

AIC – Akaike Information Criterion

ANOVA – analysis of variance

Apc – adenomatosis polyposis coli

AS – Angelman syndrome

ATR-X – X-linked alpha-thalassemia/mental retardation

BWS – Beckwith-Wiedemann syndrome

CG – cytosine and guanine

ChIP – chromatin immuno-precipitation

ChIP-seq – chromatin immuno-precipitation followed by massively parallel sequencing

COBRA – combined bisulfite and restriction analysis

CpG – cytosine followed by a guanine

CTCF – CCTC binding factor

DMD – differentially methylated domain

dn – double null

DNA – deoxyribonucleic acid

DNMT – DNA methyl transferase

DNS – de-novo signal

DT – diphtheria toxin

e – embryonic day

EB – embryoid body

EF – embryonic fibroblast cell

ES – embryonic stem cell

FISH – fluorescent *in situ* hybridization

GLM – generalized linear model

GO – gene ontology

H3K4me3 – tri-methylated histone H3, lysine 4

H3K9me3 - tri-methylated histone H3, lysine 9

H3K9ac – acetylated histone H3, lysine 9

H3K14ac - acetylated histone H3, lysine 14

H3K27me3 - tri-methylated histone H3, lysine 27

H3K36me3 - tri-methylated histone H3, lysine 36

H4K20me3 - tri-methylated histone H4, lysine 20

HDAC – histone deacetylase

HMM – hidden Markov model

ICF – immunodeficiency centromere instability and facial anomolies

ICR - imprinting control region

kb – kilo-base

Kbp – kilo-basepair

LNA – locked nucleic acid

MBD – methyl binding domain

Mbp – mega-basepairs

MEF – mouse embtyonic fibroblast

Min – multiple intestinal neoplasia

mRNA – messenger ribonucleic acid

miRNA – micro ribonucleic acid

ncRNA – noncoding ribonucleic acid

PCR – polymerase chain reaction

PGC – primordial germ cell

piRNA – piwi-associated ribonucleic acid

PRC2 – polycomb repressive complex 2

PWS – Prader-Willi syndrome

Q-PCR – quantitative polymerase chain reaction

QTL – quantitative trait loci

RdRP - ribonucleic acid dependent ribonucleic acid polymerase

RLGS – restriction landmark genome scanning

RNA – ribonucleic acid

RNAi – ribonucleic acid interference

RT-PCR – reverse transcriptase polymerase chain reaction

SNP – single nucleotide polymorphism

UTR – untranslated region

WT – wildtype

LIST OF SYMBOLS

Δ - deletion

+ - wildtype

♀- female

♂- male

I. INTRODUCTION

I.1 Epigenetic regulation of gene expression

I.1.1 Introduction to epigenetics

Historically, the term “epigenetics” has had a variety of meanings, which started out quite broad, but the definition has been refined and has become more specific over time. Epigenetics was first used by Conrad Waddington in 1942 and was a result of combining the words genetics and epigenesis, which refers to the differentiation of totipotent cells during embryonic development. As defined by Waddington, epigenetics simply referred to the study of how genes give rise to phenotypes during development (Waddington 1942). Half a century later, Robin Holliday defined epigenetics as "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms” (Holliday 1990). This definition of epigenetics applies to any non-mutational change to DNA that influences development. More recently, the emphasis on development has faded and current usage of the word refers to non-mutational changes, including DNA methylation, histone modifications, and alterations to chromatin structure, which affect gene expression. Nevertheless, controversies regarding the exact definition of epigenetics are ongoing (Ptashne 2007).

Epigenetic changes are reversible, yet can be stably inherited through cell divisions, and may last for multiple generations, while there is no change to the underlying DNA sequence. Perhaps the most basic example of epigenetic control is the process of cellular differentiation, whereby totipotent

embryonic stem cells make the decision to differentiate into the various cell types present in a mature organism. This process occurs through the activation of certain genes and the inhibition of others, which is controlled by factors other than mutational changes to DNA sequence or, epigenetic factors. The fundamental role of epigenetic mechanisms in differentiation were most recently indicated by chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) performed on cultured mouse cells representing three developmental stages (Bernstein 2006).

However, this is just one of many examples of epigenetic control. In each case, epigenetic control involves modifications to gene expression, but there are myriad distinct epigenetic phenomena that result in varying levels of gene expression. Examples include, but are not limited to, genomic imprinting, X chromosome inactivation, transvection and paramutation, position effect variegation, epigenetic reprogramming, and maternal and paternal effects. The mechanisms of epigenetic control for each of these phenomena is different, but each relies on changes to chromatin structure which can be brought about, in part, by two of the better understood elements of epigenetic control: DNA methylation and covalent histone modifications.

I.1.2 DNA methylation

In mammals, CpG dinucleotides are targets for the transfer of a methyl group from S-adenosyl methionine to carbon five of the cytosine pyrimidine ring to form 5-methylcytosine. This reaction is catalyzed by one of three active DNA methyl transferases (DNMT). DNMT3a and DNMT3b are *de novo* methyl transferases, capable of adding methyl groups to fully unmethylated DNA

strands, while DNMT1 is a maintenance DNMT that methylates hemi-methylated DNA strands. Plants also undergo DNA methylation although, in contrast to mammals, CpG, CpNpG, and CpNpNp sites can be methylated. However, yeast and worms do not exhibit DNA methylation, while *Drosophila* do so at extremely low levels (Bird 2002; Chan 2005).

DNA methylation occurs largely in repetitive elements including satellite sequences, centromeric repeats, and CpG islands in or near promoter sequences. CpG islands are CG rich stretches of DNA with high proportions of CpG dinucleotides. Although CpG islands are largely unmethylated, a significant number of CpG islands are methylated in the promoter region of genes. Promoter methylation is often associated with gene silencing events in cancer (Toyota 1999). Correspondingly, methylation of CpG islands can have a powerful effect on transcription through interaction with proteins containing methyl binding domains (MBD) (Boyes 1991). Methyl binding proteins, such as MBD2 or MeCP2, can associate with co-repressor complexes, including Sin3 or Mi2/NuRD, that are able to recruit complexes containing chromatin remodeling factors and histone deacetylases (HDAC) (Dobosy 2001). HDAC1 and HDAC2 function to remove acetyl groups from lysine residues in the N-terminal tails of histones H3 and H4. Histone deacetlyation may repress transcription as a result of an increase in the net positive charge of the histone. This, in turn, may condense chromatin by allowing a stronger association between the positively charged histone and the negatively charged DNA (Knoepfler 1999). Condensed chromatin conformations resulting from chromatin remodeling complexes or the activity of HDACs may limit transcription factor binding, resulting in transcriptional repression (Knoepfler

1999). However, the exact mechanism of repression is unknown and mechanisms other than chromatin condensation may be operating, possibly through the recruitment of the factors mentioned above.

However, not all genes are transcriptionally repressed in response to CpG methylation. At least three mouse loci, each of which is imprinted in certain tissues, have been identified that are transcriptionally activated in response to methylation at the imprinting control region: *Rasgrf1, Igf2,* and *Tsix*. When alleles of these genes are unmethylated, CCCTC binding factor (CTCF), an evolutionarily conserved 11 zinc finger protein, is able to bind and block interaction between the promoter and nearby enhancers, resulting in transcriptional silencing (Filippova 1996; Yoon 2005; Bell 2000; Chao 2002). The transcriptionally active allele at these loci is methylated at CpG dinucleotides, preventing CTCF binding and enabling transcription. The exact mechanism through which CTCF inhibits promoter-enhancer interaction is not known, but evidence suggests that it may involve the binding of Sin3A and HDACs to the zinc-finger region of CTCF, and long range interactions between DNA segments that may isolate regions into expression domains (Lutz 2000; Kurukuti 2006; Splinter 2006; Yusufzai 2004).

I.1.3 Histone modifications

In addition to DNA methylation, covalent modifications to histones can have a marked effect on gene expression. Histones are the proteins around which DNA is wrapped to form nucleosomes, and these proteins are subject to a wide variety of post-translational modifications that can enhance or repress expression of the genes to which the histones are bound. These modifications

are found on the N-terminal tails of histones and include acetylation, methylation, phosphorylation, sumoylation, ubiquitination, and ribosylation. Of these, however, only lysine acetylation and methylation have been studied in any depth. Lysine acetylation is generally associated with transcriptional activation, while lysine methylation has differing effects depending on the exact mark. For example, histone H3 lysine 9 trimethylation (H3K9me3), histone H3 lysine 27 trimethylation (H3K27me3), and histone H4 lysine 20 trimethylation (H4K20me3) are heterochromatic marks associated with gene silencing, while histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 36 trimethylation (H3K36me3) are associated with euchromatin and gene activation. Histones can interact not only with each other, but also with DNA methylation marks through various protein complexes to effect gene silencing or activation (Esteve 2006; Vire 2006; Li 2007; Ooi 2007; Smallwood 2007; Hernández-Muñoz 2005.)

Although DNA methylation and histone modification marks have historically been the focus of most studies in epigenetics, many other factors exist that play important roles in epigenetic regulation. Transcription factors directly influence the expression of the genes to which they are bound, while chromatin remodeling factors are key players in changes to chromatin structure, indirectly influencing gene expression. Noncoding RNAs are able to silence genes, either or transcriptionally or post-transcriptionally, based on sequence homology (Zaratiegui 2007). All of these factors, in combination, act to coordinate timing of gene expression and are necessary for proper growth, development, and differentiation.

1.2. Genomic imprinting

1.2.1 Introduction to genomic imprinting

Genomic imprinting is a specific epigenetic phenomenon that refers to parent-of-origin specific expression. That is, rather than expressing a given gene from both parental alleles, imprinted genes are expressed from only one parental allele. Imprinted genes are therefore referred to as being either maternally expressed or paternally expressed. Approximately 100 genes have been identified as imprinted in mice and about 60 have been identified in humans, with many of the imprinted genes organized into large imprinted clusters. These clusters often share distant regulatory elements, which control complex expression patterns along the entire cluster (Reik 2004).

Genomic imprinting is regulated by allele-specific placement of epigenetic marks and, although no unifying sequence features have been identified, several imprinting control regions (ICRs), which are responsible for controlling imprinted gene expression patterns, have been located and the epigenetic modifications at these ICRs have been characterized. Imprinted expression relies, in large part, upon differential DNA methylation to effect parent of origin specific expression patterns (Li 1993). Disruption of normal DNA methylation patterns leads to improper expression of imprinted genes and, in some cases, can result in severe phenotypic disorders. Beckwith-Wiedemann syndrome (BWS), Prader-Willi syndrome (PWS), and Angelman syndrome (AS) all result from disruptions of imprinted expression in humans (Reviewed in Robertson 2005). BWS is characterized by prenatal and postnatal overgrowth, as well as an increase in embryonic tumors. BWS results from perturbed imprinting of

either of two neighboring imprinted domains, *IGF2/H19* and *KCNQ1/KCNQ1OT1*, on human chromosome 11. PWS and AS are associated with adjacent reciprocally imprinted genes, *SNPRN* and *UBE3A*, on human chromosome 15 and are characterized by mental retardation and behavioral defects.

I.2.2 History of genomic imprinting

The first evidence for the nonequivalence of the two parental genomes came from work on mouse embryos involving both Robertsonian translocations and pronuclear transfer experiments in the early 1980s. A Robertsonian translocation results when two nonhomologous acrocentric chromosomes break at their centromeres. The long arms become attached to a single centromere and the short arms also join to form a reciprocal product, but are usually lost. This results in the inheritance of certain gene regions from only one parent and, in the case of imprinted genes within the region, can lead to physical abnormalities and intrauterine growth retardation (Engel 1980). Similarly, pronuclear transfer experiments resulting in androgenetic embryos (containing two paternal genomes) fail to undergo normal embryonic development, while gynogenetic embryos (containing two maternal genomes) show deficiencies in extraembyonic development (Barton 1984; McGrath 1984; Surani 1984). These results indicated that expression of genes necessary for normal embryonic development was not equivalent between the two parental genomes. A similar experiment involving transfer of parthenogenetic nuclei to normal enucleated eggs and normal nuclei to parthenogenetic enucleated eggs showed that this developmental defect was due to information contained in the nucleus and not the egg cytoplasm (Mann

1984). Later work with Robertsonian translocations resulting in uniparental disomies in specific regions allowed the identification of certain gene regions that show imprinted, or parent of origin specific, expression (Cattanach 1985).

The first indication that DNA methylation and genomic imprinting are linked came from a transgene study. A transgenic mouse strain carrying an autosomal transgene with elements of the RSV LTR and a translocated c-myc gene showed expression in the heart only when inherited from the father (Swain 1987). When inherited from the mother, the transgene was not expressed. This pattern of paternal allele-specific expression was correlated with parent-of-origin specific methylation of the transgene, which was observed in all tissues. Methylation of the transgene was acquired upon passage through the female germline and was erased in the male germline. This experiment provided the first direct evidence that autosomal gene expression can vary based on origin of inheritance and provided the first link between DNA methylation and imprinted expression. Shortly after this discovery, DNA methylation was shown to be responsible for parent-of-origin specific expression at two endogenous loci: *H19* and *Igf2r* (Stoger 1993; Ferguson-Smith 1993).

1.2.3 DNA methylation and genomic imprinting

The importance of DNA methylation in the control of imprinted expression was confirmed using mouse embryos lacking the maintenance DNA methyltransferase, DNMT1, which demonstrated imprinting defects (Li 1993). The differential DNA methylation mentioned above is established in a sex-specific manner during gametogenesis and is maintained throughout somatic

development. Of the known imprinted loci with defined DMDs, the majority are maternally methylated, while only three are paternally methylated: *Rasgrf1*/*A19*, *H19/Igf2*, and *Dlk1/Gtl2*.

Differential DNA methylation at the DMDs of ICRs can lead to allele-specific expression through either a noncoding RNA-mediated mechanism, the binding of methylation-sensitive enhancer blocking proteins, or a combination of these mechanisms (Mancini-Dinardo 2006; Sleutels 2002; Bell 2000; Yoon 2005). Although it is unclear whether this is the mechanism by which noncoding RNAs (ncRNAs) silence imprinted genes, ncRNAs have the ability to silence homologous messenger RNAs (mRNAs), and expression of the corresponding gene, in a sequence-specific manner. On the other hand, DNA methylation of enhancer blocking binding sites allows gene activation. DNA methylation prevents the binding of enhancer blockers which, when bound, inhibit communication between the enhancer and the promoter, resulting in gene silencing (Bell 2000; Yoon 2005).

ncRNAs are required for the allele-specific expression of several mouse imprinted loci, including *Igf2r, Slc22a2, Scl22a3,* and *Kcnq1*, and for X-chromosome inactivaton. At these loci, the DMDs are intronic, and the promoters of the ncRNA genes *Air* (at *Igf2r/Slc22a2/Slc22a3*) and *Kcnq1ot1* (at *Kcnq1*) are located within the DMDs. In both cases, the ncRNAs are expressed when the DMDs are unmethylated and are silenced when the DMDs are methylated (Thakur 2004; Wutz 1997). Allele-specific expression of these ncRNAs is necessary for the establishment of epigenetic silencing marks such as DNA methylation, and the heterochromatic histone

modifications H3K9me3 and H3K27me3 at the promoters of *Igf2r* and *Kcnq1*. Placement of these marks at the promoters results in imprinted expression of *Igf2r* and *Kcnq1* (Mancini-Dinardo 2006; Sleutels 2002). The *Air* ncRNA is also required for allele-specific silencing of the *Slc22a3* and *Slc22a2*. *Air* was recently shown to accumulate at and interact with the *Slc22a3* promoter, as well as with the H3K9 histone methyltransferase G9a in mouse placenta. Deletion of G9a results in biallelic expression of *Slc22a3*. Additionally, truncations of *Air* lead to a failure to accumulate at the *Slc22a3* promoter, reduced G9a recruitment, and biallelic expression. These results suggest that *Air*, and possibly other ncRNAs, are able to direct histone modifications and bring about allele-specific silencing at imprinted genes (Nagano 2008). In X-chromosome inactivaton, an antisense RNA known as *Xist* coats the inactive X chromosome, resulting in silencing of the entire 180Mbp X-chromosome, except for a very small number of genes which escape this inactivation (Ng 2007). Recently, PRC2 was identified as the direct target of a 1.6-kilobase ncRNA called RepA, which is located within *Xist*. Ezh2 binds the RepA ncRNA to recruit PRC2 to the inactive X. Expression of the antisense *Tsix* ncRNA inhibits this interaction on the active X. Furthermore, depletion of RepA inhibits H3K27me3 on the inactive X. Similarly, PRC2 deficiency compromises up-regulation of *Xist*. Therefore, the ncRNA RepA, together with PRC2, is required for the initiation and the spreading of X-chromosome inactivation (Zhao 2008).

The second mechanism of allele-specific expression is methylation-sensitive enhancer blocking. Binding of an enhancer blocking element leads to gene silencing by preventing communication between the enhancer and

promoter of the gene to which it is bound. One well-studied example of an enhancer blocking element involved in genomic imprinting is the zinc-finger protein, CTCF. CTCF has a series of eleven zinc-fingers that bind different sequences using varying combinations of the zinc fingers, which has made the compilation of a consensus binding site difficult. However, binding sites have been experimentally identified in several ICRs. Perhaps the best known example is the mouse *H19/Igf2* locus, which has four CTCF binding sites in the DMD (Bell 2000). On the maternal allele, which is unmethylated, CTCF is able to bind the DMD and disrupts communication between a distant enhancer and the *Igf2* promoter, preventing expression of *Igf2*, while allowing expression of *H19*. *H19* is expressed because the DMD overlaps with the *H19* promoter, and binding of CTCF allows the promoter to remain unmethylated, allowing *H19* expression. However, on the methylated paternal allele, the DMD and the *H19* and promoter are methylated and *H19* is silent. Presence of methylation at the paternal DMD blocks CTCF binding and allows the enhancer to interact with the *Igf2* promoter, allowing *Igf2* expression. If CTCF binding at this locus is disrupted, the maternal DMD gains inappropriate DNA methylation and *Igf2* is biallelically expressed, while *H19* is silenced. (Fedoriw 2004; Schoenherr 2003). While this is the best-studied example of CTCF-mediated allele-specific expression, it is not the only example. CTCF mediated allele-specific expression also occurs at the imprinted *Rasgrf1* locus, as well as in X-chromosome inactivation (Yoon 2005; Chao 2002).

Finally, there is one case where an ICR is known to effect imprinted expression by a combination of the two mechanisms discussed above. The mouse KvDMR1 ICR contains both the promoter for the *Kcnq1ot1* ncRNA and

two CTCF-binding sites that demonstrate repressive activity in enhancer-blocking assays. Deletion of the KvDMR1 results in biallelic expression of eight maternally expressed genes within the *Kcnq1* imprinted cluster, including *Cdkn1c*. While truncation of the *Kcnq1ot1* transcript results in a similar loss of imprinted expression of these genes in the placenta, it does not affect imprinted expression of *Cdkn1c* in several embryonic tissue types, despite the loss of normal DNA methylation marks at this gene. Therefore, in certain embryonic tissues, the *KvDMR1* can silence *Cdkn1c* by a mechanism independent of *Kcnq1ot1* transcription, perhaps by CTCF-associated repression. Thus far, this is the only example of an ICR silencing the same gene by both mechanisms. (Shin 2008)

I.2.4. Histone modifications and genomic imprinting

Although the role of DNA methylation in the control of imprinted expression is relatively well understood, the role of histone modifications remains largely unexplored. An intriguing result of one study that examined histone modifications in mouse imprinted genes is that the *Kcnq1/Kcnq1ot1* locus (briefly described above) is regulated by histone modifications, independent of DNA methylation, in the placenta (Lewis 2004). At this locus, the ICR is surrounded by maternally expressed imprinted genes. These genes, with the exception of *Kcnq1ot1* and *Cdkn1c*, lack allele-specific DNA methylation. Also, mutations to DNMT1 do not result in a loss of imprinting in the placenta, but do result in a complete loss of imprinting in the embryo proper. Furthermore, the H3K9 methyltransferase G9a is necessary for placenta-specific imprinting. In G9a mutants, which show a loss of both H3K9me2 and H3K9me3, imprinting in the placenta is disrupted, while the

embryo proper maintains proper imprinting (Wagschal 2008). In the placenta, the silenced paternal allele is characterized by heterochromatic histone modifications, including H3K9me2 and H3K27me3, while the active maternal allele is marked with euchromatic histone modifications, including H3K4me2 and histone H3 lysine 9 (H3K9ac) and lysine 14 acetylation (H3K14ac) (Lewis 2004; Wagschal 2008).

Additional evidence for the importance of histone modifications in imprinted gene expression comes from studies involving EED and YY1. EED is a part of the polycomb repressive complex 2 (PRC2) and YY1 is responsible for recruitment of EED. Mice with *Eed* mutations show a failure of imprinting at autosomal imprinted genes and genes involved in X-chromosome inactivation (Mager 2003). Similarly, YY1 knockdown mice show a loss of DNA methylation at the DMD of the paternally expressed gene *Peg3,* with a corresponding loss of imprinted expression (Kim 2008). Additionally, our lab has shown that H3K27me3, the placement of which is dependent on PCR2, is both mutually exclusive and mutually antagonistic with DNA methylation at *Rasgrf1*. Correspondingly, loss of YY1 increases DNA methylation at *Rasgrf1* (Lindroth 2008). While these studies covered a limited number of loci, the first studies to tackle the genome-wide relationship between histone modifications and genomic imprinting revealed a bivalent histone modification state at the ICRs of imprinted genes, where active alleles were enriched for H3K4me3 and silenced alleles were enriched for H3K9me3 (Delaval 2006; Mikkelsen 2007). Despite these advances, there is still much to discover about the role of histone modifications in imprinted gene expression, as well as how histone modifications interact with DNA methylation to coordinate genomic imprinting.

I.3 Epigenetics in development and disease

I.3.1 Developmental timing of DNA methylation marks

During early development, mammals undergo dramatic, genome-wide changes in levels of DNA methylation, which can be separated into two general categories: reprogramming of somatic cells and reprogramming of primordial germ cells (PGC). Somatic cells are reprogrammed after fertilization when both the maternal and the paternal genomes present in the zygote experience a wave of global demethylation, which removes gamete-specific methylation marks (Reviewed in Morgan 2005). This phase is completed by the blastocyst stage and is followed by a wave of *de novo* re-methylation, which establishes methylation marks important for early development (Reviewed in Morgan 2005). However, imprinted genes in somatic cells resist this wave of global demethylation and retain their parent-of-origin specific inherited methylation marks (Reviewed in Morgan 2005; Santos 2004). Defects in maintenance of DNA methylation occur at this stage, when imprinted genes fail to resist the wave of global demethylation.

PGCs undergo a separate reprogramming process. The process of reprogramming ensures that previous methylation marks are erased, and it allows the proper sex-specific methylation marks to be placed during sperm and oocyte development. Defects in establishment of DNA methylation occur at this stage, when DNA methylation marks fail to be placed during sperm or oocyte development. The PGCs of the fertilized mouse embryo maintain the inherited maternally and paternally imprinted patterns of DNA methylation until just before migration to the gonadal ridge at embryonic day 10.5 (e10.5)

(Hajkova 2002; Yamazaki 2003). Around e11.5, the PGCs arrive in the gonadal ridge, and a demethylation event occurs, removing the remaining inherited methylation patterns from the PGCs (Hajkova 2002). PGC demethylation is completed around e13-e14, when male and female PGCs enter mitotic or meiotic arrest, respectively. Maternal and paternal imprints are established before fertilization, during germ cell development, and mature sperm and oocytes each possess the correct sex-specific methylation pattern (Reviewed in Morgan 2005). Depending on the gene, methylation marks can be established either in the PGCs, in mature gametes before fertilization, or after fertilization but before syngamy, likely due to different signals for establishment of DNA methylation (Shemer 1996). The PGC reprogramming process described here is critical in ensuring proper expression of imprinted genes in subsequent generations.

I.3.2. *Trans*-acting DNA methylation reprogramming factors

As mentioned previously, there are three mammalian DNA methyltransferases, DNMT1-3, which play different roles in the establishment and maintenance of DNA methylation (Goll 2005). DNMT3a and DNMT3b, in concert with the catalytically inactive DNMT3L, are the enzymes responsible for the establishment of DNA methylation. DNMT3a and DNMT3b are highly expressed in both the male and the female germline and, in addition, they are *de novo* DNA methyltransferases (La Salle 2004). DNMT3L has no detectable DNA methyltransferase activity itself, but is necessary for the activity of DNMT3a (Hata 2002). Experiments involving mice with a germline-specific knockout of DNMT3L suggest that DNMT3a and DNMT3L physically associate and play a major role in the *de novo* methylation of paternally methylated

genes. Likewise, DNMT3a and DNMT3L are needed for the *de novo* methylation of maternally methylated genes in oocytes. However, there is no evidence that DNMT3b is needed for methylation in the female germline (Bourc'his 2004; Kaneda 2004; Lees-Murdock 2005). The only evidence thus far for the involvement of DNMT3b in the establishment of DNA methylation comes from the paternally methylated gene *Rasgrf1* (Kato 2007). As opposed to the *de novo* methyltransferases, DNMT1 is a maintenance DNA methyltransferase, which adds methyl groups to the unmethylated strand of hemi-methylated DNA, or a DNA duplex in which only one of the two strands is methylated. A specific splice variant of DNMT1, DNMT1o, is oocyte-specific and maintains imprinted DNA methylation patterns during the wave of preimplantation genome-wide DNA demethylation (Howell 2001).

In contrast to the factors that establish and maintain DNA methylation, much less is known about the factors responsible for DNA demethylation. Active DNA demethylation is thought to occur during the preimplantation wave of genome-wide DNA demethylation. The paternal pronucleus of the mouse is fully demethylated within 6-8 hours after fertilization and before the first round of DNA replication occurs (Oswald 2000). While maintenance of an unmethylated state has been attributed to factors such as CTCF and YY1 at *H19/Igf2* and *Peg3*, respectively, factors responsible for active DNA demethylation have remained a mystery until recently (Kim 2008; Donohoe 2007; Schoenherr 2003 Fedoriw 2004).

Emerging evidence suggests that DNA excision repair enzymes, similar to the repair-based DNA demethylation mechanism in *Arabidopsis*, are

involved in active DNA demethylation in vertebrates. The non-enzymatic Gadd45 proteins are important regulators of this process, and are responsible for both the recruitment of enzymatic machinery and the coupling of deamination and nucleotide-excision repair to bring about DNA demethylation (Reviewed in Ma 2009). In zebrafish embryos, cytosine methylation is removed *in vivo* through the coupling of AID, which converts 5-methylcytosine to thymine, and Mbd4, a G:T mismatch-specific thymine glycosylase. Injection of methylated DNA into zebrafish embryos induces strong DNA demethylation activity, which is absent when either AID or Gadd45 are depleted (Rai 2008). Additionally, overexpression of AID/Mbd4 *in vivo* causes global demethylation of both the genome and injected DNA fragments (Rai 2008). Finally, knockdown of either AID or Mbd4 results in remethylation of a common set of genes (Rai 2008). These results suggest that a two-step mechanism of DNA demethylation, which is promoted by Gadd45, operates in zebrafish: first, AID deaminates 5-methylcytosine, followed by thymine base excision by Mbd4.

I.3.3. *Cis*-acting DNA methylation reprogramming factors

At a handful of imprinted loci, the *cis*-acting sequence elements regulating imprinted DNA methylation have been identified. At two well-studied mouse imprinted genes, *Igf2r* and *Snrpn*, two important *cis*-acting regulatory signals have been identified. The 6-12 bp allele discriminating signal (ADS) prevents paternal allele methylation, while the 7-8 bp *de novo* signal (DNS) establishes methylation in the female germline (Wutz 1997; Birger 1999; Kantor 2004). However, these sequences have only been tested at ectopic sites in the genome. Therefore, their function at the endogenous *Igf2r* and *Snrpn* loci has not been confirmed. Studies at the endogenous

H19/Igf2 (described above) revealed that the DMD is the sequence that attracts DNA methylation on the paternal allele. This same sequence is the site where the methylation-sensitive enhancer blocking protein CTCF binds on the unmethylated maternal allele (Srivastava 2000; Thorvaldsen 1998).

Additionally, many DMDs contain or are adjacent to repetitive elements, as supported by tandem repeat microarray data (Neumann 1995; Lippman 2004; Walter 2006). As a result of this finding, and of targeted deletion experiments, tandem repeats have been implicated in the regulation of imprinted expression at mouse genes including *Rasgrf1*, *Xist* and *Tsix* (Yoon 2002; Hoki 2009; Cohen 2007). Although the mechanism through which tandem repeats effect imprinted methylation is unknown, hypotheses include siRNA mediated regulation, secondary structure formation, and germline specific repeat binding factors. Tandem repeats might produce siRNA continuously through the use of RNA-dependant RNA polymerase (RdRP) over multiple rounds of RNA interference (RNAi) (Martienssen 2003). These have been shown to regulate epigenetic phenomena in fission yeast and plants and may operate in mice as well (Alleman 2006; Chan 2004; Rassoulzadegan 2006; Verdel 2004). Secondary structure formation, such as G-quartet structures, is known to affect the ability of DNMTs to methylate the underlying DNA sequence and, although identification of germline specific binding factors has been difficult, there is evidence for protein factors, such as BORIS, that are present in only one of the two parental germlines (Smith 1994; Loukinov 2002).

Most recently, computational studies have reported sequence features

that may play a role in the regulation of imprinted DNA methylation. The relative distribution and orientation of retroposons is highly correlated with genomic imprinting (Greally 2002; Ke 2002; Luedi et al., 2005). Short interspersed nuclear elements (SINE) are depleted in the intergenic regions of imprinted genes, while long interspersed nuclear element 1 (LINE1) and endogenous retrovirus (ERV) elements are slightly enriched. These features may affect the establishment or maintenance of DNA methylation at imprinted loci during development, especially during germline development. However, it is important to keep in mind that these connections are only correlative and have not yet been shown to be causative.

I.3.4 Causes and consequences of aberrant methylation

Although DNA methylation patterns are tightly regulated throughout development under homeostatic conditions, changes in environmental stimuli can lead to disruption of normal DNA methylation, resulting in disease phenotypes (Jaenisch 2003). There are several classic examples of the effect of diet on epigenetic states. Dietary folate is a key component in one-carbon metabolism, which is necessary for the conversion of homocysteine to methionine, and eventually to S-adenosylmethionine, the primary methyl donor for DNA methylation. Different levels of folate in the diet of pregnant agouti mice can have dramatic phenotypic effects on the progeny of those mice. For example, feeding pregnant agouti mothers diets containing a range of folate levels during and after pregnancy influenced the levels of DNA methylation in their progeny, which translated to coat color and obesity phenotypes (Lillycrop 2008; Lillycrop 2007; Waterland 2003). Similarly, pregnant rats on a protein-restricted diet showed reduced DNA methylation in the promoter region of the

glucocorticoid receptor and peroxisome proliferators-activated receptor-alpha in liver (Lillycrop 2005). In humans, diseases such as cardiovascular disease and diabetes are correlated with familial nutrition, and exposure to pesticides is linked to decreases in fertility caused by altered DNA methylation in the male germline. Each of these phenotypes persists in subsequent generations (Kaati 2007; Anway 2005).

Perhaps the best-studied disease resulting from aberrant DNA methylation is cancer. The first link between DNA methylation and cancer came from genome-wide studies of methylation levels in cancer cell lines. These cells demonstrated global hypomethylation, combined with local hypermethylation (Feinberg 1983). The role of DNA methylation in tumor formation was further supported by experiments involving *Min* (multiple intestinal neoplasia) mice. These mice have a mutated form of the *Apc* (adenomatosis polyposis coli) tumor suppressor gene, resulting in high levels of colon cancer. *Min* mice containing an additional mutation in *Dnmt1,* or *Min* mice treated with 5-aza-2'-deoxycytidine, both of which result in a global reduction in DNA methylation, show a decrease in the frequency of tumor formation (Eads 2002; Laird 1995). These results provided evidence that DNA methylation can contribute to tumor formation *in vivo*.

Although most data concerning epigenetic modification and disease involves DNA methylation, defects in histone modification levels have been reporting in aging studies. *Sir2* is a NAD-dependent HDAC that has been implicated in the control of aging. In both yeast and worms, a reduction in *Sir2* expression leads to decreased lifespan, while *Sir2* overexpression lengthens

lifespan (Reviewed in Guarente 2005). As histone deacetylation is generally linked to gene silencing, it is hypothesized that *Sir2* downregulates genes involved in energy metabolism (Boily 2008). A second histone modification, H4K20me3, is also linked to aging. H4K20me3 enrichment increases with age and correlates with defects in nuclear lamina, a hallmark of aging (Sarg 2002; Shumaker 2006). However, the mechanism by which H4K20me3 contributes to laminar defects is unknown.

In addition to the phenotypes discussed above, aberrant DNA methylation patterns are responsible for the human diseases including, but not limited to, Rett syndrome, immunodeficiency centromere instability and facial anomalies (ICF) syndrome, X-linked alpha-thalassemia/mental retardation syndrome (ATR-X), and cancer (Haim 1999; Paulsen 2001; Reviewed in Plass 2002). Given the wide variety of diseases that result from improper DNA methylation patterns, understanding the mechanisms that regulate the establishment and maintenance of DNA methylation, as well as histone modifications, is a high priority. Imprinted loci provide a unique model system in which it is possible to dissect the molecular mechanisms that control both proper DNA methylation and histone modification patterns.

I.4. Identification of novel imprinted genes

I.4.1 Importance of imprinted gene identification

The identification of novel imprinted genes has become increasingly important with the realization that imprinting defects are associated with a

variety of complex disorders ranging from obesity to behavioral disorders. In addition, there are several known disorders (discussed above), which result specifically from defects in genomic imprinting of certain genes. Discovery of new imprinted genes increases the likelihood that genes underlying both complex trait and imprinting disorders will be identified. In addition, imprinted genes have the potential to aid in research regarding the regulation of DNA methylation and histone modifications. Little is known about what determines whether or not a DNA sequence is methylated or where and when a histone modification is placed. Imprinted loci are particularly desirable for studying the regulation of DNA methylation and histone modifications for two reasons: they exhibit predictable patterns of allele-specific DNA methylation and histone modifications, and corresponding patterns of allele-specific expression.

I.4.2 History of genome-wide searches for novel Imprinted genes

Despite the importance imprinted genes play in human health and the promise they hold for research on the regulation of DNA methylation and mistone modifications, only a handful of studies have tackled genome-wide identification of imprinted genes. These have done so mainly by experimental methods, with some success. One large-scale study identified candidate imprinted transcripts in the mouse genome by expression profiling of cDNA clones. cDNA microarrays were used to detect differential expression by comparing mRNA levels in the P9.5 gynogenetic and androgenetic mouse embryos (Nikaido 2003). Of the ~28,000 FANTOM2 transcripts analyzed, ~2,000 were identified as imprinted candidates. Interestingly, 39 of the 2,000 transcripts mapped to known imprinted regions of the mouse genome, while 56 were ncRNAs, and 159 were antisense transcripts. Experimental validation

of two transcripts located in the Prader-Willi syndrome region identified these transcripts as imprinted, indicating that allele-specific array-based methods are useful for large-scale identification of novel imprinted genes.

Additionally, two recent papers have successfully identified novel imprinted genes using massively-parellel sequencing approaches (Wang 2008; Babak 2008). In each case, RNA and cDNA were prepared from reciprocal F1 mouse tissues and the cDNA subjected to massively parallel sequencing. The first study used neonatal mouse brain and successfully identified 3 novel imprinted genes in this tissue. Imprinting of each gene was confirmed by Sanger and pyrosequencing of PCR products spanning allele-specific SNPs. The second used e9.5 mouse embryos from reciprocal F1 mouse crosses and identified 6 novel imprinted genes. This study also suggests that many ncRNAs are subject to imprinted expression, as more than half of all imprinted single-nucleotide polymorphisms did not overlap previously discovered imprinted transcripts and a large fraction of these represent novel ncRNAs within known imprinted loci. These studies demonstrate the feasibility of unbiased, transcriptome-wide analysis for the identification of novel imprinted genes.

I.4.3 Computational prediction of novel imprinted genes

A third method has also been used in imprinted gene identification: computational prediction. Genome-wide identification of novel imprinted genes based on sequence features alone was pioneered in a series of two studies, which used a two-tiered machine-learning program to predict novel mouse and human imprinted genes genome-wide (Luedi 2005; Luedi 2007). The first tier

used a training set of known imprinted genes and presumed non-imprinted control genes to train the prediction program based on data on a variety of sequence features, but focusing on repetitive elements and transcription factors. The second tier was where the resulting model was run on the genome-wide data to predict novel imprinted genes. And, although they did not experimentally verify any candidate imprinted genes in the mouse genome, they predicted a total of 600 imprinted mouse genes. A similar approach was used for the human genome and successfully verified two new imprinted genes on a chromosome that was not previously known to contain any imprinted genes. However, the imprinting status of genes was not verified using reciprocal F1 crosses, so false positives due to expression bias cannot be ruled out.

Given the importance of imprinted genes in human health, and the utility of imprinted genes for the study of DNA methylation, I address two main aims in this dissertation: 1) to use epigenomic and sequence features to predict novel mouse imprinted genes and 2) to experimentally verify the imprinted status of those predicted imprinted genes using material from reciprocal F1 mouse crosses.

II. SUCCESSFUL PREDICTION OF NOVEL IMPRINTED GENES FROM EPIGENOMIC DATA[1]

II.1 Abstract

Approximately 100 mouse genes undergo genomic imprinting (http://igc.otago.ac.nz/Search.html) whereby one of the two parental alleles is epigenetically silenced. Imprinted genes influence development, X-chromosome inactivation, obesity, schizophrenia, and diabetes, motivating the identification of all imprinted loci (McGrath 1984; Surani 1984; Cattanach 1985; Jiang 2004; Brown 1991; Lee 1999; Shao 2008; Xie 2008; Crespi 2008; Mackay 2008). Local sequence features have been used to predict locations of imprinted loci, but rigorous testing using reciprocal crosses validated only one, which resided in a previously identified imprinting cluster (Luedi 2005; Ruf 2007). Here we show that specific epigenetic features in mouse cells correlate with imprinting status in mice and identify hundreds of additional genes predicted to be imprinted in mouse. From this list, we identify 2 novel imprinted genes out of the first 10 genes tested. Furthermore, we use a custom microarray approach, in combination with material from reciprocal crosses, to test 563 genes for imprinted candidates. Experimental validation of 32 of these candidate genes lead to the identification of 5 additional novel imprinted genes. Our results demonstrate the utility of expanding epigenomic databases for a focused search for novel imprinted genes.

---

[1] McLean C.M., Eilertson K.E., Bustamante C.D., Soloway P.D. (2009) Successful computational prediction of novel imprinted genes. Submitted to PLoS Genetics.

II.2 Introduction

Genomic imprinting refers to genes that are expressed from one of the two parental alleles in a parent-of-origin specific manner. Thus far, about 100 mouse imprinted genes have been identified, with between 500 and 2,000 additional genes predicted to be imprinted by computational or microarray methods, respectively (Luedi 2005; Nikaido 2003). The identification of novel imprinted genes has become increasingly important with the realization that imprinting defects are associated with a variety of complex disorders such as obesity, diabetes, and schizophrenia. Similarly, several disorders exist, including Prader-Willi syndrome, Angelman syndrome and Beckwith-Wiedmann syndrome, which result specifically from defects in genomic imprinting of certain genes (Cattanach 1992; Buiting 1995; Ohlsson 1993; Lalande 1996). Discovery of new imprinted genes increases the likelihood that genes underlying both complex trait and imprinting disorders will be identified.

In addition, imprinted genes have the potential to aid in other areas of research such as regulation of DNA methylation, which plays an important role in early development and X-inactivation, as well as in genomic imprinting. Proper DNA methylation is critical, as aberrant methylation patterns can lead to numerous human diseases, including cancer. However, the regulation of DNA methylation is not well understood. Imprinted genes provide an ideal system for studying the regulation of DNA methylation for two reasons: they exhibit predictable patterns of methylation and corresponding patterns of expression.

Likewise, imprinted gene expression is controlled by allele-specific epigenetic states at imprinting control regions (ICRs). However, there is little sequence conservation among ICRs, and the DNA sequences controlling epigenetic states have been defined for only two ICRs (Shemer 2003; Yoon 2002; Kantor 2004). Therefore we reasoned that epigenomic datasets might augment the utility of sequence features to identify novel imprinted genes (Mikkelsen 2007). In support of this, characteristic epigenetic features have been identified at gene regulatory elements of both non-imprinted and imprinted genes (Heintzman 2007; Wen 2008).

Despite the importance imprinted genes play in human health, only a handful of studies have tackled genome-wide identification of imprinted genes (Wolf 2008; Pollard 2008; Schulz 2006; Smith 2003; Plass 1996; Kuroiwa 1996; Kaneko-Ishino 1995; Hayashizaki 1994; Hatada 1993; Maeda 2006). These have done so mainly by experimental methods, with some success. However, computational prediction of imprinted genes using DNA sequence features alone has recently come to the forefront (Luedi 2005; Luedi 2007). Here, we use DNA sequence and epigenomic features to identify novel imprinted genes. We first selected a set of features we anticipated might correlate with imprinting status. DNA sequence features we considered included GC content, CpG islands, miRNA clusters and predicted G-quartet sites (Huppert 2005). The genome wide epigenetic and chromatin features we considered were predicted and verified CTCF binding sites and several histone states including H3K4me3, H3K9me3, H3K27me3, H3K36me3 and H4K20me3 (http://insulatordb.utmem.edu/help.php; Mikkelsen 2007). These histone states were characterized in embryonic stem (ES), mouse embryonic

fibroblast (MEF) and neuronal progenitor (NP) cells by ChIP-chip and ChIP-seq analyses (Mikkelsen 2007).

These sequence features were selected for specific reasons. GC percentage and CpG islands were examined because the differential methylation that is associated with imprinted genes is usually placed on cytosine residues, specifically cytosine residues that are followed by guanine residues. In fact, a recent paper looked at sequence features within human, mouse, and cattle and correlated these features with imprinted genes using 20 genes known to be imprinted in all three species (Khatib 2007). miRNA clusters were included because several known imprinted gene regions are associated with miRNA clusters including the *Gtl2/Dlk1*, imprinted cluster on mouse chromosome 12 and the well studied *H19/Igf2* imprinted cluster on mouse chromosome 7 (Seitz 2004; Cai 2007). Additionally, miRNA clusters have been implicated as having a role in DNA methylation in both plants and mammals (Sinkkonen 2008). CTCF binding sites, both experimentally validated and computationally predicted, were examined because CTCF can act as a methylation sensitive enhancer blocker, leading to silencing of the allele to which it is bound. For example, CTCF is frequently associated with allele-specific silencing at imprinted genes, including *H19*/*Igf2* and *Rasgrf1* (Bell 2002; Yoon 2005). Predicted G-quartet sites were included because the secondary structures that they form can affect the ability of DNA methyltransferases, to methylate the underlying DNA sequence and imprinted expression relies heavily on DNA methylation (Smith 1994).

Finally, data on the genome wide localization of five histone

modifications in three developmental stages, mouse embryonic fibroblast cells, neural progenitor cells, and embryonic stem cells was included in the analysis (Mikkelsen 2007). The specific modification we looked at is tri-methylation of various lysine residues present in the N-terminal histone tails. The addition of methyl groups to the various lysines can have a dramatic effect on the expression of any genes to which these histones are bound. Of the five modifications examined, H3K4me and H3K36me3 are marks of active genes or euchromatin, while H3K9me3, H3K27me3, and H4K20me3 marks of repressed genes or heterochromatin. These histone states were characterized in ES, MEF and NP cells by ChIP-chip and ChIP-seq analyses. A Hidden Markov Model and sliding window method were used to identify their sites of enrichment in the genome. Additionally, we identified raw numbers of histone binding sites using raw data filtered for sites with read scores greater than two. We used each of these three data sets in our analyses. Despite the dramatic effect these marks can have on gene expression, little is understood about what controls the placement of these modifications, and we were interested to know whether they are associated with imprinted genes. Especially considering that specific histone modifications, overlapping H3K4me3 and H3K9me3 in ICRs, have been correlated with imprinting status in the past and that at least one histone modification is know to have an antagonistic relationship with DNA methylation (Mikkelsen 2007; Lindroth 2008).

A generalized linear model (GLM), along with a training array of 53 known imprinted genes and 84 non-imprinted genes, was used to select sequence features that aid in prediction of imprinted status within each of 11 gene regions (100, 10, and 1 kb upstream of genes, within genes, 5'UTRs,

exons, introns, 3'UTRs, 1, 10, and 100 kb downstream of genes). Those genes that were predicted to be imprinted by 5 or more of the 11 models were initially considered candidate imprinted genes. A subset of 10 of the imprinted candidates was subjected to experimental validation, yielding 2 novel imprinted genes. Based on this success, we expanded our candidate list to genes that were predicted to be imprinted by 3 or more of the 11 models. We then used a microarray analysis to narrow our list of candidate genes. A subset of these genes was tested for imprinted expression by RT-PCR on F1 progeny from reciprocal crosses of polymorphic mouse strains, followed by either restriction enzyme digestion or Sanger sequencing. This approach yielded an additional 5 novel imprinted genes.

II.3 Results

II.3.1 Histone modifications strongly correlate with imprinting status

To use DNA sequence and epigenomic features for identifying novel imprinted genes, we first selected a set of features we anticipated might correlate with imprinting status. DNA sequence features we considered included GC content, CpG islands, miRNA clusters and predicted G-quartet sites (Huppert 2005). The genome-wide epigenetic and chromatin features we considered were predicted and verified CTCF binding sites (http://insulatordb.utmem.edu/help.php), because of their known importance in controlling imprinted gene expression, and several histone states including H3K4me3, H3K9me3, H3K27me3, H3K36me3 and H4K20me3, which collectively influence gene expression and cellular differentiation. These

histone states were characterized in ES, MEF, and NP cells by ChIP-chip and ChIP-seq analyses (Mikkelsen 2007). We used two data sets which report regions enriched for histone modfifications as identified by either a Hidden Markov Model (HMM) or a sliding window method (WIN,) as originally reported (Mikkelsen 2007). We also included a data set that reported the raw number of histone modification binding sites throughout the genone, as identified by ChIP-seq positions that had a read score of two or greater. There were a total of 29 sequence and epigenetic features.

Having identified epigenomic and sequence features that might be useful for identifying novel imprinted genes, we then used [R] (http://www.r-project.org/) to determine the extent to which each of the epigenomic and sequence features correlated with imprinting status. For this test, we used a set of 53 known imprinted genes (Table II.S1), comparing them to 29,544 non-redundant mouse transcripts. For each gene, we compared features present in 11 different regions relative to the transcription start site: 100kb upstream, 10kb upstream, 1kb upstream, within genes, 5'UTRs, exons, introns, 3'UTRs, 1kb downstream, 10kb downstream, and 100kb downstream. From the correlation analysis, a pattern of histone modifications associated with imprinted genes emerged (Figure II.1). In general, the repressive histone modifications H3K9me3, H3K27me3, or H4K20me3 tended to be associated with imprinted genes in any of the three cells types (embryonic stem cells, embryonic fibroblast cells, or neural progenitor cells) and over nearly all of the 11 domains. Furthermore, H3K36me3, a histone modification enriched in active chromatin, was associated with non-imprinted genes in the region 10kb downstream of genes. The positive association between H3K9me3,

Table II.S1. Genes included in the training set. A total of 137 genes, including known imprinted and non-imprinted genes, were included in the training data set. Fifty-three genes fall within the known imprinted category, while 84 genes fall within the non-imprinted category. We compiled the 53 known imprinted genes from the Imprinted Gene Catalogue (http://igc.otago.ac.nz/Search.html). The non-imprinted genes were assumed to be non-imprinted based on the lethality of homozygous mutations, and the viability of heterozygous mutations in mice as described in the Jackson Laboratories MGI database (http://www.informatics.jax.org/).

Known Imprinted Genes		**Non-Imprinted Genes**		
Apoc2	*Osbpl5*	*Acvrl1*	*Fn1*	*Nog*
Asb4	*Peg10*	*Ada*	*Foxd1*	*Nte*
Calcr	*Peg12*	*Adcy3*	*Foxg1*	*Ntrk1*
Cd81	*Peg3*	*Adcy7*	*Fst*	*Ntrk2*
Cdkn1c	*Phlda2*	*Akp2*	*Gba*	*Ntrk3*
Commd1	*Plagl1*	*Apaf1*	*Gdf1*	*Otx1*
Copg2	*Pon2*	*Arnt*	*Gja1*	*Ppard*
Dcn	*Pon3*	*Ascl1*	*Hbb*	*Psen1*
Dlk1	*Ppp1r9a*	*Ass1*	*Hdh*	*Rela*
Gatm	*Rasgrf1*	*Atoh1*	*Hgs*	*Runx1*
Gnas	*Rb1*	*Bcl2*	*Inhba*	*Shh*
Grb10	*Sgce*	*Bdnf*	*Itga5*	*Slc4a1*
Gtl2	*Slc22a18*	*Cdh1*	*Itgav*	*Smo*
H19	*Slc22a2*	*Cdh2*	*Itgb1*	*Snap25*
Htr2a	*Slc22a3*	*Cdk5*	*Jun*	*Sp4*
Igf2	*Slc38a4*	*Chrna3*	*Jup*	*Tal1*
Igf2r	*Snrpn*	*Col3a1*	*Kcnj2*	*Tgfb1*
Impact	*Snurf*	*Csk*	*Klf1*	*Tgfbr1*
Ins2	*Tnfrsf23*	*Cxcr4*	*Kras*	*Tsc2*
Kcnq1	*Tspan32*	*Cycs*	*Lifr*	*Unc5c*
Magel2	*Tssc4*	*Dgat2*	*Mad2l1*	*Vcam1*
Mash2	*U2af1-rs1*	*Dnmt1*	*Maf*	*Vhlh*
Mcts2	*Ube3a*	*Edn3*	*Map2k4*	*Wnt3a*
Mest	*Zim1*	*Ednrb*	*Mdm2*	*Wnt5a*
Mkrn3		*Efemp2*	*Mgat1*	*Wnt7b*
Nap1l4		*En1*	*Myb*	*Wt1*
Nap1l5		*Epas1*	*Myf5*	
Ndn		*Evi1*	*Myf6*	
Nnat		*F5*	*Nf1*	

Figure II.1. Correlation of features in each region with imprinting. For the 11 gene regions examined, the correlation coefficients were calculated for the 29 features using the cor() function in [R]. P-values for correlation coefficients were calculated using a two-tailed t-test and were considered significant if less than 0.000157 after Bonferroni correction for the 319 comparisons (values in Table II.S2). Log transformed p-values were calculated and depicted in a heat map, where a log transformed p-value less than -8.76 is significant. The significance level of the log-transformed p-values is indicated by a color gradient, with colors at the red end of the gradient representing higher levels of significance. Features are clustered according to the dendrogram on the left according to similarities in p-value distributions. The cell sources for each of the histone modifications data sets is indicated (ES, embryonic stem cells; EF, embryonic fibroblast cells; NP, neural progenitor cells), as is the method used to calculate histone modification enrichment (HMM, Hidden Markov Model; WIN, sliding window model) (Mikkelsen 2007).

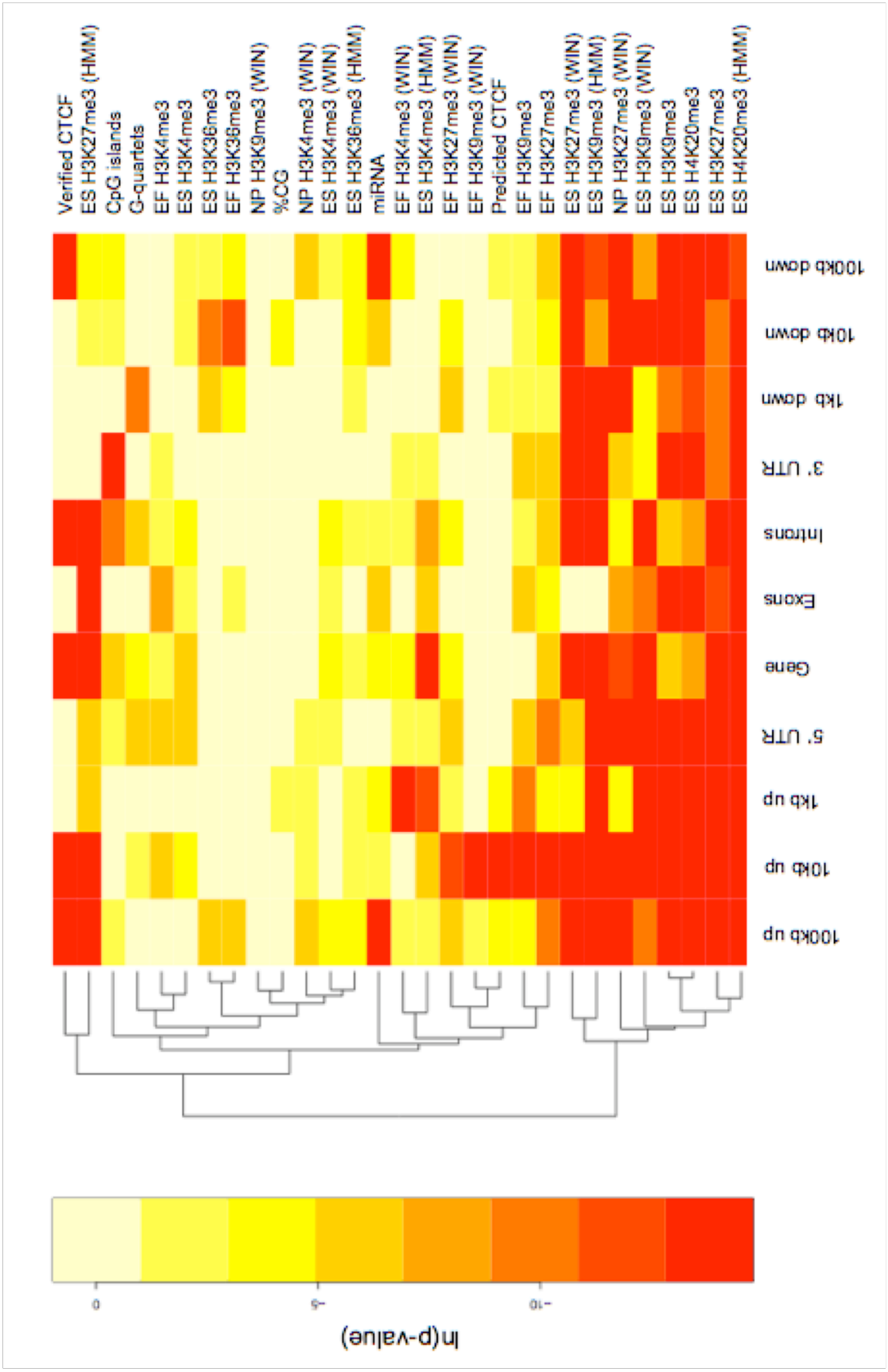

Verified CTCF
ES H3K27me3 (HMM)
CpG islands
G-quartets
EF H3K4me3
ES H3K4me3
ES H3K36me3
EF H3K36me3
NP H3K9me3 (WIN)
%CG
NP H3K4me3 (WIN)
ES H3K4me3 (WIN)
ES H3K36me3 (HMM)
miRNA
EF H3K4me3 (WIN)
ES H3K4me3 (HMM)
EF H3K27me3 (WIN)
EF H3K9me3 (WIN)
Predicted CTCF
EF H3K9me3
EF H3K27me3
ES H3K27me3 (WIN)
ES H3K9me3 (HMM)
NP H3K27me3 (WIN)
ES H3K9me3 (WIN)
ES H3K9me3
ES H4K20me3
ES H3K27me3
ES H4K20me3 (HMM)
100kb up
10kb up
1kb up
5' UTR
Gene
Exons
Introns
3' UTR
1kb down
10kb down
100kb down
0
-5
-10
ln(p-value)

Table II.S2. Correlation coefficient p-values for features enriched in imprinted genes by gene region used to generate Figure II.1. For each gene region examined, the correlation coefficient p-value indicating the significance of correlation with imprinting status for each of the features of interest included in our analysis is shown. Correlation coefficients were calculated using the cor() function in [R]. P-values for the correlation coefficients were calculated using a two-tailed t-test and were considered significant if less than 0.000157 (0.05 p-value/319 comparisons) after Bonferroni correcting for multiple comparisons. Features with significant p-values are highlighted in either green, to signify a correlation in the positive direction, or red, to signify a correlation in the negative direction. The columns labelled “Pos” and “Neg” tally the number of regions in which each feature correlates with imprinted genes in a positive and a negative direction, respectively. ES stands for embryonic stem cells, EF stands for embryonic fibroblast cells, and NP stands for neural progenitor cells. HMM stands for enrichment as determined using a Hidden Markov Model, while WIN stands for enrichment as determined using a sliding window model (Mikkelsen 2007).

Feature	100up	10up	1up	5'UTR	In Gene	Exons	Introns	3'UTR	1dn	10dn	100dn	Pos	Neg
%CG	0.912378	0.590616	0.043953	0.345161	0.287104	0.260332	0.813265	0.434307	0.294456	0.007114	0.200820	0	0
G-quartets	0.703111	0.133142	0.680119	0.001767	0.009755	0.968707	0.004456	0.768283	0.000098	0.757766	0.445824	1	0
miRNA	0.000000	0.050574	0.006366	0.859144	0.014490	0.003085	0.040250	0.818769	0.820586	0.003972	0.000000	2	0
CpG island	0.044549	0.773639	0.811157	0.042787	0.004542	0.704237	0.000046	0.000000	0.368262	0.044848	0.029007	2	0
Verified CTCF	0.000000	0.000000	0.859144	0.952835	0.000000	0.875636	0.000000	0.933337	0.933337	0.812930	0.000000	5	0
Predicted CTCF	0.013586	0.000000	0.026743	0.625201	0.683401	0.587129	0.681047	0.591622	0.064826	0.233170	0.077739	1	0
EF H3K4me3	0.548611	0.003186	0.548387	0.001679	0.058327	0.000727	0.034300	0.165958	0.220383	0.440995	0.868248	0	0
EF H3K9me3	0.006238	0.000002	0.000100	0.005080	0.372446	0.002976	0.171234	0.005558	0.139872	0.063811	0.053517	2	0
EF H3K27me3	0.000072	0.000000	0.007393	0.000062	0.003116	0.019873	0.002924	0.004269	0.101935	0.006277	0.003064	3	0
EF H3K36me3	0.001142	0.811425	0.714943	0.763260	0.621611	0.176142	0.766424	0.329263	0.006740	0.000020	0.018163	0	1
ES H4K20me3	0.000000	0.000000	0.000000	0.000000	0.000878	0.000001	0.000633	0.000000	0.000024	0.000000	0.000005	9	0
ES H3K4me3	0.336046	0.020766	0.254961	0.001146	0.005397	0.097354	0.005807	0.987056	0.227799	0.046610	0.146762	0	0
ES H3K9me3	0.000000	0.000000	0.000000	0.000000	0.001267	0.000001	0.001114	0.000000	0.000061	0.000000	0.000000	9	0
ES H3K27me3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000006	0.000000	0.000096	0.000084	0.000070	0.000000	11	0
ES H3K36me3	0.005576	0.868870	0.595748	0.490950	0.424074	0.178215	0.633940	0.510597	0.004685	0.000147	0.102057	0	1
ES H3K4me3 (HMM)	0.171486	0.004015	0.000013	0.039469	0.000005	0.005329	0.000423	0.138784	0.388726	0.353818	0.251803	2	0
ES H3K9me3 (HMM)	0.000000	0.000000	0.000000	0.000000	0.000000	0.577023	0.000000	0.000000	0.000000	0.000794	0.000007	9	0
ES H3K27me3 (HMM)	0.000000	0.000000	0.002089	0.002150	0.000000	0.000000	0.000000	0.529679	0.366802	0.046649	0.017285	5	0
ES H3K36me3 (HMM)	0.007396	0.107181	0.039114	0.401464	0.073591	0.388030	0.041782	0.426347	0.100102	0.007915	0.013052	0	0
ES H4K20me3 (HMM)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000006	11	0
ES H3K4me3 (WIN)	0.006011	0.341020	0.296925	0.161729	0.015081	0.062799	0.014067	0.565580	0.989288	0.216237	0.036394	0	0
ES H3K9me3 (WIN)	0.000047	0.000000	0.000000	0.000000	0.000000	0.000176	0.000000	0.006699	0.011259	0.000000	0.000197	7	0
ES H3K27me3 (WIN)	0.000000	0.000000	0.012421	0.002909	0.000000	0.504756	0.000000	0.000000	0.000000	0.000000	0.000001	8	0
EF H3K4me3 (WIN)	0.126426	0.341020	0.000000	0.042801	0.020089	0.866204	0.078138	0.114392	0.436092	0.443941	0.018221	1	0
EF H3K9me3 (WIN)	0.072684	0.000002	0.701288	0.824815	0.326421	0.220057	0.597690	0.762867	0.725864	0.254941	0.552456	1	0
EF H3K27me3 (WIN)	0.002882	0.000013	0.116783	0.003378	0.008004	0.297751	0.020161	0.983317	0.002083	0.018956	0.353656	1	0
NP H3K4me3 (WIN)	0.001221	0.061094	0.096342	0.045705	0.984355	0.834144	0.772224	0.881972	0.789637	0.552109	0.004411	1	0
NP H3K9me3 (WIN)	0.522721	0.646058	0.831093	0.884796	0.386553	0.771782	0.409352	0.771914	0.767341	0.690569	0.259141	0	0
NP H3K27me3 (WIN)	0.000000	0.000000	0.012421	0.000000	0.000007	0.000325	0.006452	0.004969	0.000000	0.000000	0.000000	7	0

H3K27me3 and imprinting is in close agreement with their documented role in mechanisms controlling imprinted DNA methylation at *Rasgrf1*. At that locus, H3K27me3 excludes DNA methylation from the unmethylated maternal allele and H3K9me3 is needed for optimal placement of DNA methylation on the methylated paternal allele (Lindroth 2008). Because the associations we found among H3K9me3, H3K27me3, and imprinting were known to be mechanistically relevant, this provided a measure of confidence in our approach. Additional positive correlates included predicted G-quartet sites, miRNA clusters, and verified CTCF binding sites.

II.3.2 Histone modifications are important predictors of imprinted status

With significant correlations among imprinting status and sequence and epigenetic features identified, we then used the features present in the 11 domains to train a set of GLMs. To do this, we used a training set of 53 known imprinted genes, 84 likely non-imprinted genes and [R] to fit a logistic regression model for each of the 11 domain-based data sets. The response for our model was whether or not the gene is imprinted, and the potential predictors were the set of 29 sequence and epigenetic features. The predictors included in the model were chosen using stepwise selection based on the Akaike Information Criterion (AIC).

The predictors included in each of the 11 resulting logistic regression models varied from region to region. The model significance level assigned by [R] for each of the features of interest in each gene region examined is shown in Table II.1. From this modelling, several features of interest stand out as being effective predictors of imprinting. Six histone features were predictive of

Table II.1. Significance levels of features included in prediction models by region. A total of 137 genes, including known imprinted and non-imprinted genes, were included in the training data set. Fifty-three genes fall within the known imprinted category, while 84 genes fall within the non-imprinted category. We compiled the 53 known imprinted genes from the Imprinted Gene Catalogue (http://igc.otago.ac.nz/Search.html). The non-imprinted genes were assumed to be non-imprinted based on the lethality of homozygous mutations, and the viability of heterozygous mutations in mice as described in the Jackson Laboratories MGI database (http://www.informatics.jax.org/).

Feature	100up	10up	1up	5'UTR	In Gene	Exons	Introns	3'UTR	1dn	10dn	100dn	***	**	*	+	Total
%CG	***	**	*							**	*	1	2	2	0	5
G-quartets			+					*	*			0	0	2	1	3
miRNA	+	+			+					+	+	0	0	0	5	5
CpG island	*		**		*			+	+	+		0	1	2	3	6
Verified CTCF	+										*	0	0	1	1	2
Predicted CTCF					**		**		+		*	0	2	1	1	4
EF H3K4me3	*	***										1	0	1	0	2
EF H3K9me3		+			*							0	0	1	1	2
EF H3K27me3	+					***			+	*	***	2	0	1	2	5
EF H3K36me3		*	**			***		**	***	***	*	3	2	2	0	7
ES H4K20me3		**	*	+		**			*	*		0	2	3	1	6
ES H3K4me3			**	**	+	*					*	0	2	2	1	5
ES H3K9me3				*		**			*	**	*	0	2	3	0	5
ES H3K27me3	***	***				*		**	**	***	***	4	2	1	0	7
ES H3K36me3	***	*			+			*			**	1	1	2	1	5
ES H3K4me3 (HMM)	+	*	***				+	*		*	*	1	0	4	2	7
ES H3K9me3 (HMM)	+							+	**	+	*	0	1	1	3	5
ES H3K27me3 (HMM)	+	+		**		+	*	+	+			0	1	1	5	7
ES H3K36me3 (HMM)	+	*	+		**		***	*			**	0	3	2	2	7
ES H4K20me3 (HMM)	***	+	+		*	+	**		+	+	***	2	1	1	5	9
ES H3K4me3 (WIN)		*				*	*	**				0	1	3	0	4
ES H3K9me3 (WIN)			+		**		*					0	1	1	1	3
ES H3K27me3 (WIN)					*		*	+		*	***	1	0	3	1	5
EF H3K4me3 (WIN)			*		+							0	0	1	1	2
EF H3K9me3 (WIN)	*	+										0	0	1	1	2
EF H3K27me3 (WIN)	*			+				+				0	0	1	2	3
NP H3K4me3 (WIN)	*					+						0	0	1	1	2
NP H3K9me3 (WIN)	+											0	0	0	1	1
NP H3K27me3 (WIN)	**	*								*	*	0	1	3	0	4

imprinted status in at least seven of the 11 GLMs: ES H4K20me3 (HMM) in nine of 11 models; and EF H3K36me3, ES H3K27me3, ES H3K4me3 (HMM),ES H3K27me3 (HMM), and ES H3K36me3 (HMM) in seven of 11 models. The most included non-histone features are CpG islands clusters, used in six of the 11 prediction models, and %CG and miRNA clusters, used in five of the 11 prediction models. Likewise, the features that were identified as being highly significant in the greatest number of prediction models are the histone modifications ES H3K27me3 and EF H3K36me3.

II.3.3 Model sensitivity and specificity

To assess the effectiveness of our prediction method, we analyzed a separate set of 29 genes using our 11 logistic regression models. This set included 9 known imprinted and 20 likely non-imprinted genes, none of which were included in our training data set (Table II.S3). From our analysis of the test set, we determined the sensitivity of our models, calculated as the number of known imprinted genes in the test data set that were correctly identified as imprinted; and their specificity, calculated as the number of non-imprinted genes in the test data set that were correctly identified as non-imprinted. Within each model, we identified genes as predicted to be imprinted if p, the probability the gene is imprinted, was greater than or equal to 0.8.

When we used stringent prediction criteria, requiring that five or more models predict a gene to be imprinted, sensitivity was 66.7% (six of 9 known imprinted genes were called as imprinted), while specificity was 100% (none of 20 non-imprinted genes were called as imprinted). The six genes correctly identified in the training data set were *Air, Ddc, Inpp5f_v2, Peg10, Sfmbt2,* and

Table II.S3. Genes included in the test data set. A total of 29 genes, including known imprinted genes and non-imprinted genes, were included in the test data set. 9 genes fall within the known imprinted category, while 20 genes fall within the non-imprinted category. Genes in bold were called as imprinted at the stringency levels indicated in the text.

Known Imprinted	Non-Imprinted	
Air	*Alx4*	*Ntf3*
Atp10a	*Bub1*	*Phoxa2*
Ddc	*Casr*	*Rxrn*
Dhcr7	*Cpt1a*	*Sall2*
Gabrb3	*Eln*	*Ski1*
Inpp5f_v2	*Gja5*	*Sod2*
Sfmbt2	*Hoxc13*	*Tbx4*
Tfpi2	*Inppl1*	*Trp35b2*
Th	*Kcna2*	*Vcl*
	Myh6	*Wrn*

Th. The sensitivity did not increase when we used less stringent criteria, requiring only three or more models to predict imprinting, although the specificity dropped to 95%; one of 20 non-imprinted genes, *Myh6*, was incorrectly predicted to be imprinted gene, assuming our criteria for identifying non-imprinted genes were valid. If the stringency was increased, requiring 6 or more models predict a gene to be imprinted, the sensitivity fell to 33.3% (three of 9 known imprinted genes were called as imprinted). Since no sensitivity was gained with stringency lower than prediction by five or more models, we used this stringency level for our genome wide analysis.

II.3.4 Genome-wide prediction of imprinted status

For the genome wide analysis, we used our 11 models to query 29,544 non-redundant mouse transcripts for their predicted imprinting status. When genes predicted as imprinted by five or more models are considered, we identified a candidate list of 155 genes (Table II.2). We were concerned that this list of genes was biased by the presence of an adjacent known imprinted gene, whose ICR may act over a large genomic domain. This did not appear to be the case as there was at least one gene that was predicted as non-imprinted between our 155 candidates and known imprinted genes. Reassuringly, four of the five known mammalian microimprinted genes were correctly predicted as imprinted: *Nap1l5, Nnat1, Inpp5f_v2,* and *U2af1rs1* and the imprinting status of three of the host genes in which the microimprinted genes are located was correctly classified as well (Table II.S4). In the two instances where the host gene was incorrectly classified, they were classified as non-imprinted.

Table II.2 Genes predicted as being imprinted by five or more models. Genes predicted as being imprinted by five or more models are listed. Next to the column containing the gene names is a column indicating the number of prediction models that classify each gene as being imprinted. Gene names in bold indicate that gene was experimentally tested for imprinted expression. The asterisk next to the known placental imprinted gene *Th* indicates that it was tested despite previous evidence of maternal expression (Schulz 2006).

Gene	# Models	Gene	# Models	Gene	# Models	Gene	# Models
Gpa33	10	*LOC629678*	6	*AK146072*	5	*Lgi1*	5
6430706D22Rik	10	*LOC633640*	6	*AK147510*	5	*LOC432436*	5
Dusp27	9	***Nef3***	6	*AK154031*	5	*LOC432823*	5
Mid1	9	*Neurod2*	6	*AK158329*	5	*Mam13*	5
BC096660	8	*Odz4*	6	*Amph*	5	*Metapl1*	5
Rpo1-4	8	*Ugt1a10*	6	*B230363K08Rik*	5	*Mst1r*	5
Th*	8	*Ugt1a2*	6	*BC006684*	5	*Neu4*	5
9530015I07Rik	7	*Ugt1a5*	6	*BC007165*	5	*Nfasc*	5
A530088H08Rik	7	*Ugt1a6a*	6	*BC054080*	5	*Nhlrc1*	5
AK016821	7	*Ugt1a6b*	6	*Bcl11a*	5	*Oasl1*	5
AK046026	7	*Ugt1a7c*	6	*Bsnd*	5	*Otx2*	5
BC005471	7	*Ugt1a9*	6	*Cd244*	5	*Oxsm*	5
BC099500	7	*Umodl1*	6	*Cd59b*	5	*Pag1*	5
C030011J08Rik	7	*Vil2*	6	***Cdh13***	5	*Parva*	5
Hoxc13	7	*Vnn3*	6	*Cflar*	5	*Pcgf4*	5
Slc38a1	7	*Wdr27*	6	*Chrnb3*	5	*Pigl*	5
Sytl3	7	*Zfp11*	6	***Cntn3***	5	*Pitx2*	5
Ugt1a1	7	*Zfp286*	6	*Cryba2*	5	*Piwil1*	5
Zfp629	7	*Zmat4*	6	*Cyp4f15*	5	*Pkp1*	5
1700029J07Rik	6	*3110070M22Rik*	5	*Dennd1a*	5	*Pnpt1*	5
4930478A21Rik	6	*4921537P18Rik*	5	*Dspp*	5	*Prss35*	5
9130017C17Rik	6	*4930417M19Rik*	5	*Enpp3*	5	*Ptf1a*	5
A830031A19Rik	6	*4930539E08Rik*	5	*Fbxo40*	5	*Ptprn2*	5
AK143924	6	*4933403G14Rik*	5	*Fyco1*	5	*Rgs8*	5
Ank1	6	*5330420D20Rik*	5	*Gm889*	5	*Rnf36*	5
BC046305	6	*6430573F11Rik*	5	***Gnao1***	5	*Scin*	5
BC048950	6	*A030007L22*	5	*Gng2*	5	*Shcbp1*	5
BC065085	6	*A630008I04*	5	*Grhl2*	5	*Slc6a17*	5
BC089472	6	*AB125595*	5	*Hlcs*	5	*Sspn*	5
Bcmo1	6	*Adam18*	5	*Hoxa10*	5	***Syt9***	5
Cdh15	6	*Agrn*	5	*Hoxb9*	5	*Tcam1*	5
Cmah	6	*AK004563*	5	*Hoxc4*	5	*Tmprss2*	5
Cyp2j13	6	*AK016553*	5	*Hps3*	5	*Trpc2*	5
Dynlt1	6	*AK029828*	5	*Hunk*	5	*Ube2l6*	5
Edar	6	*AK052253*	5	*Ifitm7*	5	*Xkr5*	5
Gm944	6	*AK131836*	5	*Irf8*	5	*Zfp160*	5
Hmox2	6	*AK133237*	5	*Kcnmb4*	5	*Zfp180*	5
Il3ra	6	*AK138193*	5	*Kit*	5	*Zfp445*	5
Kcnq2	6	*AK140614*	5	*Lair1*	5	*Zfp706*	5

Table II.S4. Summary of microimprinted genes and hosts. Each of the five known microimprinted genes and their host genes is listed, along with the expression status of each host gene. Also indicated is the expression status predicted by our models for each gene and its host.

Microimprinted Gene	Microimprinted Gene Prediction	Host Gene	Host Gene Status	Host Gene Prediction
Inpp5f_v2	Imprinted	*Inpp5f*	Non-imprinted	Non-imprinted
Nap1l5	Imprinted	*Herc3*	Non-imprinted	Non-imprinted
Nnat1	Imprinted	*Blcap*	Imprinted	Non-imprinted
Peg13	Non-imprinted	*1810044AZ24Rik*	Imprinted	Non-imprinted
U2af1rs1	Imprinted	*Commd1*	Imprinted	Imprinted

II.3.5 Experimental testing reveals two novel maternally expressed genes

We then tested a subset of the 155 candidate genes for experimental evidence of imprinting. Genes were selected for experimental validation using a variety of criteria. First, we selected genes that were called as imprinted by eight or more of the eleven prediction models. Second, since many imprinted genes occur in clusters, we examined whether any of the genes predicted as being imprinted by five or more models fell within 1Mb of known imprinted genes. Third, we noted whether any of the genes predicted as being imprinted by five or more fell into gene ontology (GO) categories that are significantly overrepresented in imprinted genes (Figure II.S1). Finally, genes that were not only expressed, but were expressed at high levels in either brain or placenta were added to the list. Within this list, we identified genes containing SNPs between the AKR/J and PWD/PhJ mouse strains. A total of ten out of the 155 candidate genes were ultimately selected for experimental validation (Table II.2, genes in bold font). In addition, *Th* was predicted to be imprinted, and was selected for experimental validation, but experimental data demonstrating imprinting appeared in the Otago database after we assembled our sequence data for the 29,544 non-redundant mouse transcripts (Schulz 2006). However, our results confirm previous reports of imprinted maternal expression in placenta (Figure II.2C).

The 10 genes selected for experimental validation were tested for evidence of imprinting in embryonic day 17.5 (e17.5) mouse placenta (Table II.1, genes in bold font). For each test, we used F1 crosses between the polymorphic mouse strains AKR and PWD. Importantly, we analysed

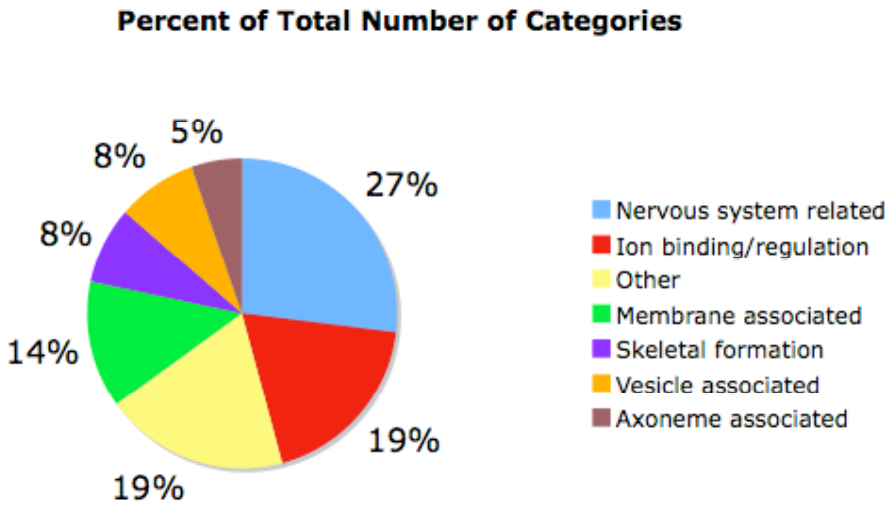


Figure II.S1. Gene ontology categories overrepresented in both known imprinted genes and predicted imprinted genes. To identify any trends in function among imprinted genes, known imprinted genes included in the training array were analyzed with GOEAST (http://omicslab.genetics.ac.cn/GOEAST/php/batch_genes.php), which identified gene ontology (GO) classes that are statistically overrepresented in imprinted genes. 209 GO classes are overrepresented in known imprinted genes ($p<0.05$) and 103 GO classes are overrepresented in the candidate imprinted genes ($p<0.05$). 37 categories that are overrepresented in the candidates gene set are similar to categories overrepresented in the known imprinted gene set. Within the 37 similar GO categories, several trends appear: 27.0% of the 37 similar GO categories are nervous system related, 18.9% are implicated in ion binding/regulation, and 13.5% are membrane related.

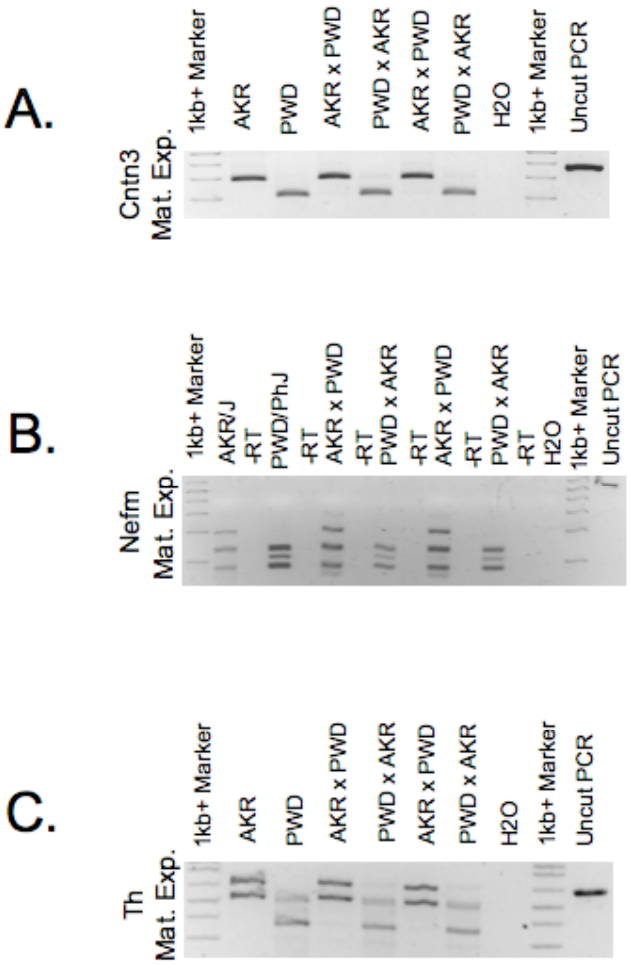


Figure II.2. Maternal allele-specific expression of two novel imprinted genes. We amplified placental cDNA from e17.5 embryos prepared using reciprocal crosses between PWK and AKR mice. PCR primers were specific to the candidate genes *Cntn3, Nefm,* and *Th* (A, B, C). PCR products were digested overnight with restriction enzymes specific for one parental allele and run on 3% agarose gels. *Cntn3*, *Nefm,* and *Th* show expression patterns consistent with maternal allele expression.

materials from reciprocal crosses to control for strain-specific expression QTL. This differs from previous attempts to validate predicted imprinted genes in human, which could not use reciprocal F1 crosses to assess imprinting status and hence could not distinguish imprinting from expression QTL. Primers were designed to span SNPs within allele-specific restriction sites (Table II.S5). These genes are not likely to be regulated by ICRs in known imprinting domains; the imprinted gene closest to *Cntn3* on chromosome 6 is 44Mb away (*Nap1l5*) and the imprinted gene closest to *Nefm* on chromosome 14 is 7Mb away (*Htr2a*, http://igc.otago.ac.nz/home.html).

To rule out false positives due to maternal tissue contamination, we assayed our samples for allele specific expression of one known paternally expressed gene: *Magel2*. *Magel2* is expressed in the maternal decidua (Figure II.3). Our samples showed expression was exclusively from the paternal allele of *Magel2*, indicating that maternal tissue contamination was negligible (Figure II.3). None of these 10, or 6 additional, genes tested were imprinted in postnatal day 2 mouse brain (not shown).

II.3.6 Microarray validation of 563 candidate imprinted genes

After seeing that the computational prediction method greatly increased our ability to detect novel imprinted genes, we designed a custom microarray to experimentally test a larger number of candidate genes for evidence of imprinting. The genes used in this approach were predicted as imprinted by three or more of the prediction models, which translated to ~1,300 genes. For each gene, 3' biased SNPs were identified between the AKR and PWD mouse strains. Three sets of probes were designed around each SNP. One set

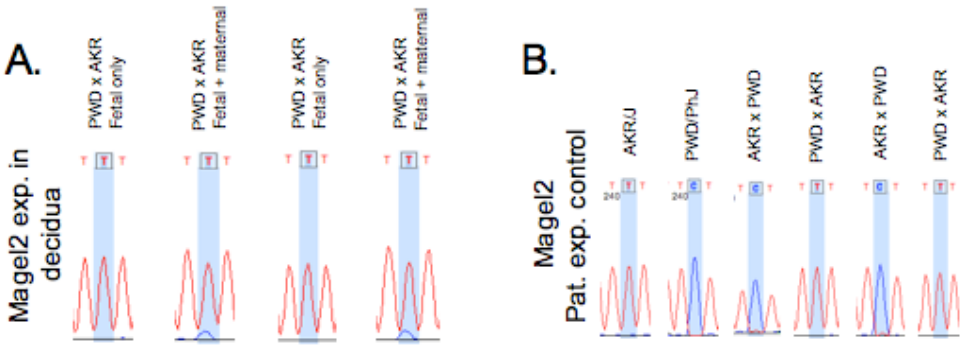


Figure II.3 Maternal tissue contamination is negligible. Placenta from e17.5 embryos were dissected to either retain or eliminate the maternal decidua. Tissue was subjected to RNA isolation followed by cDNA synthesis and RT-PCR using primers specific to *Magel2* and spanning a SNP between the strains AKR and PWD. Available SNPs for *Magel2* did not overlap with allele-specific restriction enzyme sites, so *Magel2* PCR products were gel-purified (Qiaex Quick Spin, Qiagen) and sequenced. In this cross, the maternal strain is PWD (P) and the paternal strain is AKR (A). In panel A, the samples containing only fetally derived portions of the placenta show no evidence of expression from the maternal PWD allele. However, in the samples containing both the fetally derived portion of the placenta as well as the maternal decidua, expression from the maternal PWD allele is evident, indicating that *Magel2* is expressed in the maternal decidua. In panel B, Imprinting analysis of *Magel2* using reciprocal F1 crosses demonstrates expression exclusively from the paternal allele, indicating that maternal tissue contamination is negligible.

contained the SNP position directly in the center of the probe, with all four bases represented at the SNP position. The second set contained the SNP position one base upstream from the center of the probe, with all four bases represented at the SNP position. The second set contained the SNP position one base upstream from the center of the probe, with all four bases represented at the SNP position, for a total of 12 probes per SNP. When possible, we used multiple SNPs per gene.

Each slide contained 8 identical arrays, to which were hybridized two biological replicates each of e17.5 brain and placenta samples from AKR/PWD reciprocal crosses, for a total of one sample per array. After hybridization, one-way ANOVA analysis was performed to compare the normalized fluorescent intensities of the four nucleotides at each SNP position. Those genes demonstrating reciprocal monoallelic expression of the expected SNP nucleotides were considered for further analysis. Of the candidate lists of 64 genes in placenta and 40 genes in brain, 32 and 8 genes, respectively, were tested by PCR followed by allele-specific restriction digestion or Sanger sequencing. This analysis yielded 5 novel placental imprinted genes and no novel brain imprinted genes (Figure II.4). Again, these genes are not likely to be regulated by known ICRs, as they are all located greater than 3Mb from known imprinted clusters and one gene, *Klrb1f*, is located on chromosome 13, a chromosome not previously known to contain imprinted genes. This gave us further assurance that our model was not biased by nearby known imprinted genes and was capable of identifying novel imprinted genes outside of known imprinted clusters.

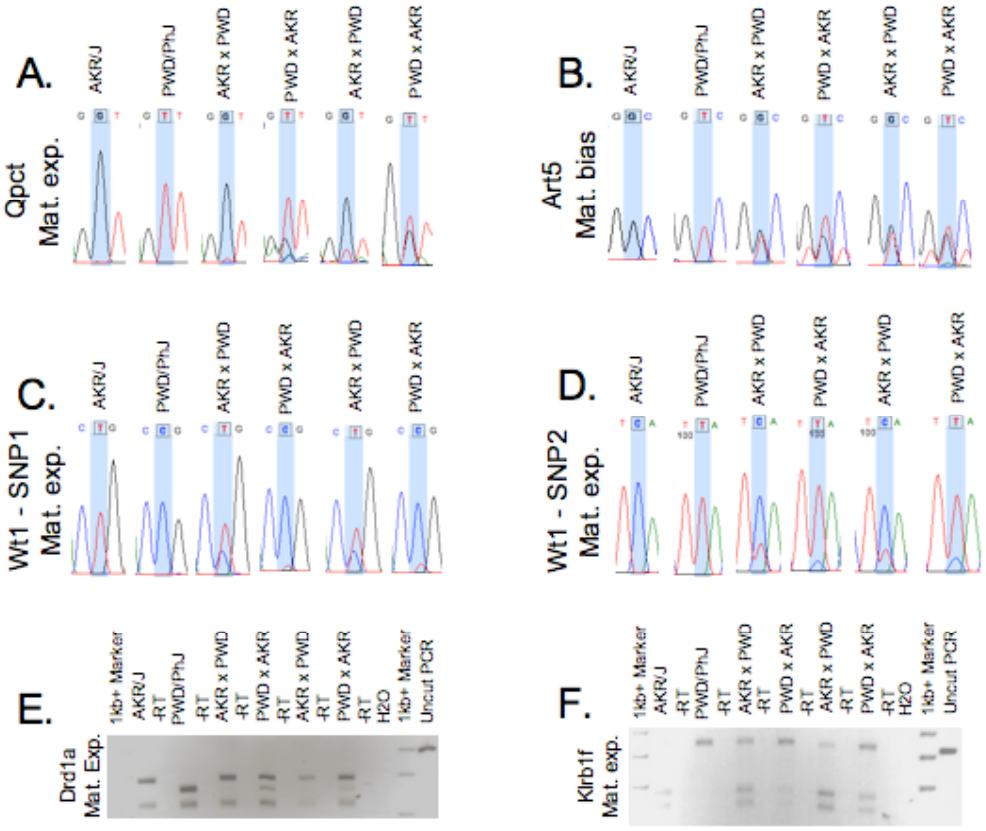


Figure II.2. Maternal allele-specific expression of four novel imprinted genes. We amplified placental cDNA from e17.5 embryos prepared using reciprocal crosses between PWK and AKR mice. PCR primers were specific to the candidate genes *Qpct, Art5, Wt1, Drd1a* and *Klfb1f* (A, B, C, D, E, F). Available SNPs for *Qpct, Wt1,* and *Art5* did not overlap with allele-specific restriction enzyme sites, so these PCR products were gel-purified (Qiaex Quick Spin, Qiagen) and sequenced. PCR products for *Drd1a* and *Klrb1f* were digested overnight with restriction enzymes specific for one parental allele and run on 3% agarose gels. All 5 genes show expression patterns consistent with maternal allele expression.

Table II.S5. Primers used for allele-specific expression assays. Primers for each gene experimentally examined for evidence of imprinted expression are listed. For each primer pair that was used in conjunction with an allele-specific restriction enzyme, the PCR band size before and after digestion is given. Information regarding whether primers span an intron, the applicable SNP and corresponding polymorphic mouse strains, and restriction enzyme used (where applicable) are noted as well. Genes highlighted in grey indicate either primer pairs that did not amplify or genes that, upon experimental testing, did not actually contain a SNP between the two mouse stains used. The presence of "(c)" next to a gene name designates a known imprinted control gene. To peform RT-PCR, RNA was subjected to random primed reverse transcription to make cDNA, and was PCR amplified using the primers indicated. Standard PCR was run with GoTaq DNA polymerase (Promega) for 40 cycles (95°C for 30 s, 55°C for 30 s and 72°C for 50 s) followed by a final extension of 5 min at 72°C. The resulting PCR products (300-700bp) were digested with an allele-specific restriction enzyme to determine parent of origin specific allelic expression patterns. PCR products were sequenced to verify amplification specificity.

Gene	-RT	Tissue Tested	MGI SNP ID	RE	Cross Needed	PCR (bp)	Expected Bands (bp)	Forward (5' - 3')	Reverse (5' - 3')
6030405A18Rik	YES	P	rs30523543	ClaI	AKR/PWD	226	74/152 vs. 226	CCCGAATGACAAGTCAACCT	TTAGTGGCGATGCCTCTTTT
9430015G10Rik	YES	P	rs33232729	AluI	AKR/PWD	315	152/163 vs. 315	GCCACTGGCAGTTTATGGAT	CCATCCCCAGAGACACACTT
9530015I07Rik	NO	n/a	rs50181318	HpyCH4III	AKRJ/PWD	435	196/249 vs. 435	GGAAAATTGGCAAGACCAGA	GTGTGCATCCTCATGTGTCC
A830018L16Rik	YES	P	rs32537969	n/a	AKR/PWD	278	n/a	CAACTGCCCCGATTTGATTA	TTCATTAAAGCCCACCAAAAA
AK016821	NO	n/a	rs37897697	NlaIII	AKRJ/PWD	408	147/261 vs. 408	TGAGAAACCCCAGACTTGGA	TCTCGCTTGCACAAGACTGT
AK046026	YES	n/a	rs49964787	BsoBI	AKRJ/PWD	521	132/389 vs. 521	CCCAACACAATAGCTGCTGA	AAGGTGAAAAGGGCAGGACT
Arhgap18	YES	P	rs39149866	Tsp45I	AKR/PWD	396	175/221 vs. 396	CCATCGTCAAGACCCAGACT	CAAAAAGCCCTGGGGATAAT
Arhgef3	YES	P	rs6214108	n/a	AKR/PWD	317	83/108/126 vs. 108/209	TCGGCCACTCCTCTTTTCTA	TTTGAACCCATCCTTGCTCT
Art5	NO	P	rs31839713	n/a	AKR/PWD	374	n/a	AGCCACTTTAACTGCGCCTA	GCCATCCAGTCTTTACCTTGC
Ccr9	YES	P	rs48048570	n/a	AKR/PWD	386	n/a	GGATCTGGTGAAGACCCTGA	AGCATCCTCCTCAGAACTGC
Cdh13	NO	B, P	rs37221381	Tsp45I	AKRJ/PWD	416	9/112/295 vs. 9/407	GTGCTCCTGGTCACGTCTG	GGCTGTCTCTGGTTCTCTGG
Cflar	YES	P	rs13468974	n/a	AKR/PWD	296	30/33/56/177 vs. 30/33/233	GTTGTGTTTGCAGGGCCTAT	TGTAATGACTCCTTAAAATGAAACAGA
Cntn3	NO	B	rs13478935	PciI	AKRJ/PWD	365	160/204 vs. 365	GGATGAGAAGCACTGTGTCG	TTCCAACCTCACAGTTGAACC
Cntn3	NO	P	rs13478935	NlaIII	AKRJ/PWD	365	78/79/208 vs. 78/208	GGATGAGAAGCACTGTGTCG	TTCCAACCTCACAGTTGAACC
Cpm	YES	P	rs48284033	n/a	AKR/PWD	291	n/a	GACTTCATGCGGGTTTTTGT	AAGCTCAGAGGTAGTCATGGTC
Dlk1 (c)	YES	P	rs50424874	n/a	AKRJ/PWD	385	n/a	CCATCTGCTTCACCATCCTG	CGTTCACTCGATTCCACACA
Drd1a	YES	B, P	rs48196064	Cac8I	AKR/PWD	303	n/a	AAATAATTGGCCCTCCCTGT	CGATGAGGCACAGCTCATTA
Dusp27	NO	B, P	rs31389162	BstUI	AKRJ/PWD	458	167/291 vs. 458	AACCGAAAGCATCTTCATGG	TCCACCTCTGGGAAGTCATC
Eif2b5	NO	P	rs48607872	HpyCH4III	AKR/PWD	390	70/75/245 vs. 145/245	GCTGGTTCAGCCAAAGAGAC	AGGGAAAGCTGGAACACAAA
Entpd1	YES	P	rs49894562	n/a	AKR/PWD	243	113/130 vs. 243	GGCCCTCTGCTCTTAACCTC	CTGCCTCCCCTCACTAAACA
Entpd1	YES	P	rs47465876	n/a	AKR/PWD	220	11/11/24/86/88 vs. 11/11/86/112	ATGTGCTCGTGCAGCTAATG	TTTTTCCAATTCTGGGTGGT
Gnao1	YES	B, P	rs32048898	DdeI	AKRJ/PWD	555	57/197/301 vs. 555	GTTTTCCTGGGACAGCACAT	TTCTAGCCTGGTCTGCCAGT
Hjurp	NO	B, P	rs47242786	BsmAI	AKRJ/PWD	359	155/204 vs. 359	CACGTACGCCATGGAGTCTA	GCTGCCTAGGATCACCAGAG
Hjurp	NO	B, P	rs47242786	BsmAI	AKRJ/PWD	431	119/312 vs.. 119,155,157	CACGTACGCCATGGAGTCTA	ACGGCTTACCCATGATTGTC
Hjurp	YES	B	rs30298180	Hpy188III	AKRJ/PWD	336	16/320 vs. 16/132/204	CCCATGGCTAGAAATCTGGA	CACCCTTTGGCCTCTAAGTG
Hmgcs2	YES	P	rs36506234	n/a	AKR/PWD	337	6/19/52/80/65/115 vs. 6/19/52/80/180	TGTTGAACCTTGAGGGAAGC	CCAACAATGGGACACAGAAA
Hs6st1	YES	P	rs31399915	n/a	AKR/PWD	351	n/a	ATGGGGACTTTCCCATTCTC	AGACAGAGGCAGGGAGTGTG
Isx	YES	P	rs13479838	RsaI	AKR/PWD	372	47/94/231 vs. 47/325	ATGACCCACAGTCCAGGAAG	TCAGCTAACCTGGGCAGAAT
Itsn1	YES	P	rs33601143	BstUI	AKR/PWD	393	75/318 vs. 393	CTGAGAAGGGAAAACCCACA	GTGGAATGTTGCAGAGCAGA
Kcnj15	YES	P	rs47548183	PleI	AKR/PWD	228	40/188 vs. 228	AGACAGCTTTCTGATGGCAAC	TCAGAGGTGCGTCTCTCCTT
Kcnq2	NO	B, P	rs27665342	HinFI	AKRJ/PWD	663	243/420 vs. 143/243/277	CTGGAGCTGCTGAGGAATCT	GCGGTCCTTATCCGTTATTG
Kdr	YES	P	rs33606899	n/a	AKR/PWD	365	n/a	TGTGTTGAGGATGGGTGAGA	AGCAAGCTGCATCATTTCCT
Klrb1f	YES	P	rs36918569	TaqI	AKR/PWD	326	148/178 vs. 326	TTGACCCTATCCCCACTTCA	AGGACATCCCACTTTCATGC
Laptm4b	YES	P	rs13460434	BsrGI	AKR/PWD	303	35/86/182 vs. 86/217	TCACACTCCCTTTTCCTTGG	TGCAGACAAACTTTAATTGCAAAG
Magel2 (c)	YES	B, P	rs33055130	n/a	AKR/PWD	330	n/a	GGGACATTGTGCTTTCTGGT	TTTTCTCCAGACAGTATTTTACCG
Mpv17l	YES	B	rs37484918	n/a	AKR/PWD	520	410/110 vs. 520	CCTCCTGCCTCTGCATTCTA	GCTGAAATGATCCTCCTCCA
Mrs2l	YES	B	rs37484918	HaeII	AKR/PWD	371	183/188 vs. 371	CCTTGGAAGACAGCTGGAAG	GCGAGGAAGTGAGGTCTGTG
Nef3	YES	B, P	rs30748232	BstNI	AKRJ/PWD	480	221/259 vs. 480	AAATCCCCTATGCCCAAATC	AGTGACCACCCCTTTCTCCT
Nef3	YES	P	rs31130948	AluI	AKR/PWD	616	81/82/85/107/127/134 vs. 81/82/127/134/192	AGTGGTGGTCACCAAGAAGG	TGTGCAGATCAGCAAGGTTC
Nlgn1	YES	P	rs33775315	n/a	AKR/PWD	369	71/298 vs. 369	CAGCTGGAGGGTTTGTCTTT	TCTTGAAAAAGGGTAGCTGGA
Pcdh21	YES	P	rs51706664	n/a	AKR/PWD	320	n/a	ATGATGAGGTTGGGAACAGC	AAATCCTCTGCTTTGCAGGA
Peg3 (c)	YES	P	rs45778768	HinFI	AKRJ/PWD	429	11/22/2643/75/84168 vs. 11/22/26/75/84/211	GGATGCACTGATGGGAAACT	GCTGGCAAGCTATGGATGAT
Ptprn2	NO	P	rs51240328	Sau96I	AKR/PWD	238	32/60/63/83 vs. 32/60/146	AGCACTTGAGGGACCAGAGA	CTAGCCACCTGCACTGACTG
Qpct	YES	P	rs52229296	n/a	AKR/PWD	320	n/a	TGATGCCTTCACTAATTTTCCTC	CATGAAAATAAATTCACAATATGCAA
Rasd2	YES	n/a	rs31959150	n/a	AKR/PWD	326	32/79/215 vs. 31/32/79/184	CAGGAAAGGTTGGAAGGACA	GACCTGGGCCTTGAGAGAAT
Rasgrf1 (c)	NO	B, P	rs29947965	HinFI	AKRJ/PWD	355	59/134/162 vs. 59/296	GGCTCATGATGAATGCCTTT	TACAGAAGCTTGGCGTTGTG
Rnf36	NO	n/a	rs27425655	MspA1I	AKRJ/PWD	419	178/241 vs. 419	CAAAGCCAAGTTGCCTGACT	AGGTTCTCTCCATGCTCTGG
Rpo1-4	NO	B	rs30703724	BsmAI	AKRJ/PWD	553	63/475 vs. 63,115,360	GCCTGCACTGATCAGCAGTAT	GCTGGAGGTAGGCAGTGAAG
Slc38a1	NO	B, P	rs49654319	MspA1I	AKRJ/PWD	416	192/224 vs. 416	ATGAGAAGCTGGGAGAGCAA	AAGGTGACATACTTTGGTGTGC
Stk4	YES	P	rs27300226	n/a	AKR/PWD	399	n/a	GCAGTGGACTCTCTGCTTCC	TGCAGGATCTGCAGTGATGT
Syt9	NO	B, P	rs49223695	Hpy188III	AKRJ/PWD	624	16/66/241/301 vs. 16/301/307	CACAGAGCCAACACCATCAG	GCCGTTGGAAGATAGCAGAG
Tbl1xr1	YES	B	rs36665902	n/a	AKR/PWD	326	n/a	AGAACATCCCAAGGCACATC	ATAACCTCCCTGCAATGGTG
Th	NO	B	rs33824309	Hpy99I	AKRJ/PWD	634	132/502 vs. 132/159/343	TGAAGCCAAAATCCACCACT	GGCATGACGGATGTACTGTG
Tinagl	NO	P	rs27517674	HpaII	AKR/PWD	397	88/134/175 vs. 134/263	AGTACCGCAGACATGGGACT	AGCCTGGTGCATCTTTGTCT
Tmed4	NO	B	rs26899778	AclI	AKR/PWD	369	181/188 vs. 369	GGTTCTCCAAGAAGCAGACG	GTCTGGACTGGTGGTTTTCG
Tmem169	YES	P	rs36357460	StyI	AKR/PWD	318	151/167 vs. 318	AGGAAGGGGAACAGGGATTA	CTGACCTCTTGGGACTCAGC
Trappc9	NO	P	rs31410905	NdeI	AKR/PWD	342	75/267 vs. 342	CGGTGTGCACAACTATGACC	AATAGGCCAGAAGCAAAGCA
Trim25	YES	B, P	rs27099824	n/a	AKR/PWD	293	98/130/163 vs. 163/228	ATTCATTGTACAGGGCAAGG	AGCAGGCTTATCCCTCATCA
Ubxd2	YES	P	rs13464870	RsaI	AKR/PWD	241	17/105/119 vs. 17/224	CTCAGGTCCAGGCTCAAGTC	GCCCAGCTCCATGTGTACTT
Ugt1a1	YES	n/a	rs30305269	BsrDI	AKRJ/PWD	411	180/231 vs. 411	TGAAAGTCAAAAGGGATTCCA	GGCTCAGAAGATCCTGGACA
Wt1	YES	P	rs27444810	n/a	AKR/PWD	336	n/a	TGAAATTCCTCCCACTCACC	ACACATGCCCTGGCCTATAA
Wt1	YES	P	rs27444811	n/a	AKR/PWD	336	n/a	TGAAATTCCTCCCACTCACC	ACACATGCCCTGGCCTATAA
Zfp629	YES	B, P	rs33162562	HaeII	AKRJ/PWD	697	270/427 vs. 132/270/295	GGCAAGAGCTTTAGCCAGTG	TGGCACTCAGTGTGAAGGTC

II.4 Materials and Methods

II.4.1 Sequence data collection

Genome-wide information on the location of CpG islands, and miRNA clusters within the Known Genes track was downloaded from the UCSC Genome Browser website (http://genome.ucsc.edu/cgi-bin/hgGateway, Feb. 2006 build), using the table feature. The location of all non-redundant known genes, exons, introns, 5' UTRs and 3' UTRs was obtained in the same fashion and the start and end position of the regions 1, 10, and 100kb upstream of each gene, as well as 1, 10, and 100kb downstream were calculated from this information. Data on the locations of experimentally verified and computationally predicted CTCF binding sites within the mouse genome were obtained from IndulatorDB (http://insulatordb.utmem.edu/).

Genome-wide data on the distribution of the histone modifications H3K4me3, H3K9me3, H3K27me3, H3K36me3, and H4K20me3 was downloaded from (ftp://ftp.broad.mit.edu/pub/papers/chipseq/) for two developmental stages: embryonic stem cells and embryonic fibroblasts. The raw data were filtered to include only sites with a read score of two or greater. Data on sites enriched for histone modifications, either calculated by a sliding window method or a hidden Markov model, were also downloaded for embryonic stem cells, embryonic fibroblast cells, and neural progenitor cells.

The entire mouse genome was downloaded from the UCSC Genome Browser website (http://genome.ucsc.edu/cgi-bin/hgGateway, Feb. 2006 build) and QUADPARSER (http://www-shankar.ch.cam.ac.uk/quadparser.html,

Huppert 2005) was used to scan the genome for sites with potential to form G-quartet structures. A custom Perl script was designed to calculate the %GC within each of the gene regions (gene, exon, intron, 5' UTR, 3'UTR, +1kb, +10kb, +100kb, -1kb, -10kb, and -100kb) for each known gene.

For each of the known genes, custom Perl scripts were used to tally the number of times each of the features of interest occurred within each of the 11 gene regions (gene, exon, intron, 5' UTR, 3'UTR, +1kb, +10kb, +100kb, -1kb, -10kb, and -100kb).

Microarray expression data for each of the 155 genes predicted as being imprinted by 5 or more prediction models was obtained from the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway, Feb. 2006 build).

II.4.2 Calculation of correlation coefficients and p-values

Correlation coefficients were calculated using the cor() function in R (http://www.r-project.org/). P-values for the correlation coefficients were calculated using a two-tailed t-test and were considered significant if less than 0.000157 (0.05 p-value/319 comparisons) after Bonferroni correcting for multiple comparisons.

II.4.3 Training data set

A list of 53 mouse imprinted genes was compiled from the Imprinted Gene Catalog (http://cancer.otago.ac.nz/IGC/Web/home.html) to use in the training array (Table II.S1). An additional 84 non-imprinted genes was

included in the training array (Table II.S1). The non-imprinted genes used in our training data set were identified in a systematic search of the Jackson Laboratories MGI database (http://www.informatics.jax.org/) for mice with knockout mutations demonstrating homozygous lethality. We selected those genes that were viable as heterozygotes, but that were lethal as homozygotes.

II.4.4 Training procedure

Using the glm() command in R we fit a logistic regression model for each of the 11 data sets using the training data. The response for our model is whether or not the gene is imprinted, and the potential predictors were the set of 29 features. The predictors to be included in the model were chosen using stepwise selection based on the Akaike Information Criterion (AIC). The predictors included in each of the 11 the resulting logistic regression models varied from region to region (Table II.1). The resulting 11 models were then each tested on a corresponding gene region data set where the genes included were known to be imprinted or non-imprinted.

II.4.5. Test data set

A list of 9 mouse imprinted genes was compiled from the Imprinted Gene Catalog (http://cancer.otago.ac.nz/IGC/Web/home.html) to use in the training array (Table II.S2). An additional 20 non-imprinted genes was included in the training array (Table II.S2). Non-imprinted genes used in our training data set were identified in a systematic search of the Jackson Laboratories MGI database (http://www.informatics.jax.org/) for mice with knockout mutations demonstrating homozygous lethality. We selected those genes that were viable as heterozygotes, but that were lethal as homozygotes.

II.4.6 Candidate list compilation

To obtain a list of candidate imprinted genes to subject to the first round of experimental validation, the predicted imprinted genes identified in each gene region were intersected. Only genes predicted to be imprinted by 5 or more models were considered in the initial candidate list. This list was narrowed further by considering proximity to known imprinted genes (within 1Mb), number of models predicting each gene to be imprinted, GO classification, and expression levels.

II.4.7 Tissue collection and RNA preparation for confirmation of imprinting status

AKR/J and PWD/PhJ reciprocal timed matings were performed by placing males and females together overnight and checking for evidence of a copulatory plug in the morning. Pregnant females were sacrificed at day 17.5 of pregnancy. Whole brain and placenta (lacking the decidua) were dissected out and snap frozen in liquid nitrogen. Total RNA was extracted from the F1 brain and placenta samples using the guanidium thiocyanate method. For each sample, the RNA concentration and the A260 nm/A280 nm ratio was checked using a NanoDrop ND-1000 Spectrophotometer.

II.4.8 SNP and restriction site identification

SNPs were identified in the final candidate imprinted genes using the Jackson Laboratory Mouse Genome Informatics website. In each case, potential SNPs and 10bp to either side was analyzed using NEB cutter (http://tools.neb.com/NEBcutter2/index.php) to select SNPs that overlap with a

restriction site for identification of allele-specific expression.

II.4.9 Confirmation of imprinting by RT-PCR and digestion or sequencing

For each gene, primers were designed using Primer3 (http://fokker.wi.mit.edu/primer3/input.htm) to overlap an allele-specific restriction site (Table II.S5). Where possible, primers were placed to span introns. To perform RT-PCR, 5ug of RNA from reciprocal crosses between polymorphic mouse strains (B6xCast, or PWDxAKR) was subjected to random primed reverse transcription to make cDNA, and was PCR amplified using the primers in Table II.S5. Standard PCR was run with GoTaq DNA polymerase (Promega) for 40 cycles (95°C for 30 sec, 60°C for 30 sec and 72°C for 50 sec) followed by a final extension of 5 min at 72°C. The resulting PCR products (300-700bp) were either digested with an allele-specific restriction enzyme, or Sanger sequenced, to determine parent of origin specific allelic expression patterns. All PCR products were sequenced to confirm amplification specificity (Figure II.S2)

II.4.10 Identification of SNPs for array-based allele-specific expression analysis

Genes predicted to be imprinted by three or more models (~1,300 genes) were considered in the initial candidate list. SNPs from mouse strains PWD and AKR were identified for each of these candidate-imprinted genes in the Jackson Laboratories MGI database (http://www.informatics.jax.org/). Only genes containing SNPs in the 3'UTR, or within 1kb of the 3' end of the gene, were included for experimental validation, narrowing the list to 563 genes. Where possible, multiple SNPs were used for each gene.

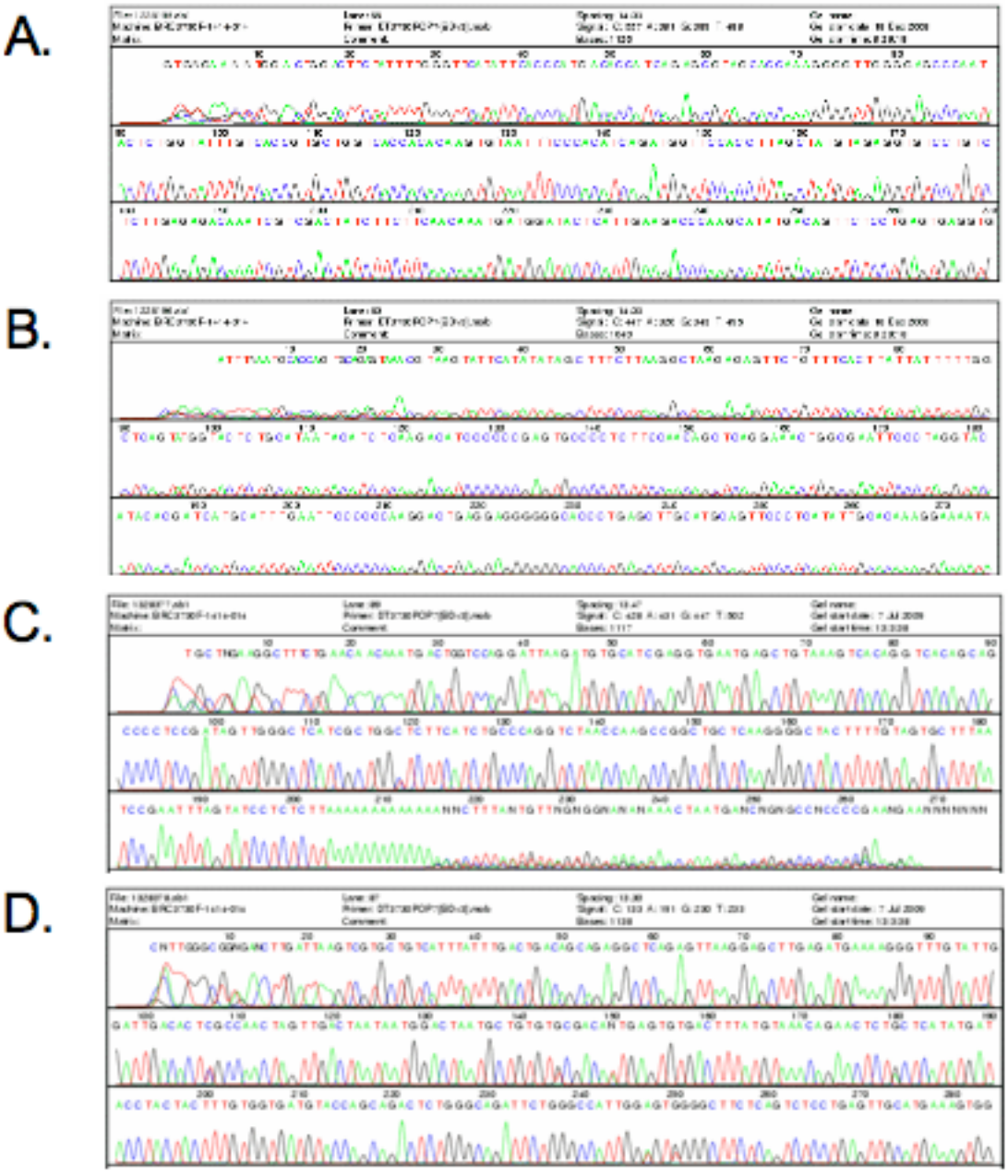


Figure II.S2. Sequence trace files of *Cntn3, Nefm, Drd1a,* and *Klrb1f* PCR products. Sequence trace files of gel-purified PCR products from primers amplifying *Cntn3* (A), *Nefm* (B), *Drd1a* (C) and *Klrb1f* (D). Sequencing results confirm that the expected products were amplified and that products are specific to the expected genes.

II.4.11 Microarray probe design

50bp of sequence to either side of each usable SNP was collected. Using this information, 12 microarray probes were designed for each SNP. Four probes contain the SNP centered within the probe and have the bases A, T, C, or G at the SNP position. Four probes contain the SNP 1bp upstream from the center of the probe and have the bases A, T, C, or G at the SNP position. Four probes contain the SNP 1bp downstream of the center of the probe and have the bases A, T, C, or G at the SNP position. All 12 probes for each SNP were trimmed to the same length, which was determined by melting temperature. For each SNP, all 12 probes are between 25 and 31bp and the length was chosen so that the melting temperature of each of the 12 probes was above 50C. Probe quality was analyzed using Agilent's e-array website (http://www. earray.chem.agilent.com/).

II.4.12 Tissue collection and RNA preparation for microarray hybridization

AKR/J and PWD/PhJ reciprocal timed matings were performed by placing males and females together overnight and checking for evidence of a copulatory plug in the morning. Pregnant females were sacrificed at day 17.5 of pregnancy. Whole brain and placenta (lacking the decidua) were dissected out and snap frozen in liquid nitrogen. Total RNA was extracted from two biological replicates of the F1 brain and placenta samples using the Qiagen RNeasy Lipid Tissue Mini Kit. For each sample, the RNA concentration and the A260 nm/A280 nm ratio was checked using a NanoDrop ND-1000 Spectrophotometer. RNA quality was determined using the Agilent 2100 Bioanalyzer. All of the samples hybridized to the microarray had a RNA

integrity number between 9.7 and 10.

II.4.13 Microarray experiment

RNA was subjected to oligo dT primed cRNA amplification. cRNA was synthesized using cyanine-3 labeled CTP and was hybridized to a custom Agilent 8x15K array by the Cornell Microarray Core facility. Hybridization was carried out at 50C. In both cases, material from reciprocal crosses was included to rule out false positived from expression QTLs. Material hybridized to the array consisted of two biological replicates of each reciprocal cross each from e17.5 brain and e17.5 placenta.

II.4.14 Data analysis and candidate imprinted gene identification

After hybridization, the microarray slide was scanned with the Axon 4000B scanner and normalized fluorescence intensities were calculated using GenePix Pro 6.0 software and the background subtraction method. One-way ANOVA analysis was performed to compare the averaged normalized fluorescence intensities of the four nucleotides at each SNP position. In total, the ANOVA test was done twelve times: once for each of the three probe sets, using materials from two reciprocal crosses and two tissue types. If the fluorescence intensities for a gene were found to be unequal, the level of the most highly expressed nucleotide was compared to that of the second most highly expressed nucleotide to determine if there was a significant difference. Candidate imprinted genes were identified as those genes demonstrating reciprocal monoallelic expression of the expected SNP nucleotides. From this list, a subset was selected for further examination based on whether the normalized fluorescence intensity of the most highly expressed nucleotide was

significantly higher than that of the second most highly expressed nucleotide. A subset of 32 of the 64 placenta candidates, and all 8 of the brain candidates, was subjected to experimental validation as described in II.4.9. These genes were selected for experimental validation over others because the results from multiple probe sets were in agreement for these genes.

II.5 Discussion

Genome-wide identification of novel imprinted genes based on sequence features alone was pioneered by a series of two papers by Luedi *et. al*. The first paper of the set predicted a total of 600 imprinted genes in the mouse genome (Luedi 2005). This prediction was done using data on a variety of sequence features, but focusing on repetitive elements and transcription factors, which were run through a two-tiered machine-learning program. The first tier used a training set of known imprinted genes and presumed non-imprinted control genes to train the prediction program. The presumed non-imprinted genes used in this data set were comprised of 500 genes that were chosen at random from the genome and assumed to be non-imprinted, whereas the genes we used as non-imprinted controls were selected in a search for mutations that, while lethal in the homozygous state, show no difference in phenotype between heterozygous and wildtype mice. This allowed an extra level of confidence in our non-imprinted control genes. Also, the known imprinted genes used by Luedi *et al*. were not necessarily known to be imprinted in mouse. For some, the imprinting status was only known in human. In the training array described here, all of the known imprinted genes used were known to be imprinted in mice, as there are

documented cases of genes being imprinted in humans and not in mice and *vice versa*.

As with our method, the second tier was where the resulting model was run on the genome-wide data to predict novel imprinted genes. And, although they did not experimentally verify any candidate imprinted genes in the mouse genome, they did the in the human genome and successfully verified two new imprinted genes on a chromosome that was not previously known to contain any imprinted genes (Luedi 2007). Furthermore, these were the only two genes they tested experimentally.

Here we have discussed a similar computer-based approach to the prediction of novel mouse imprinted genes. One difference between our approach and the approach discussed above is the modeling method used. The GLM method that we chose is desirable for modeling biological data sets because of its flexibility. For example, GLMs can model both linear and nonlinear relationships between various types of response variables and predictor variables. Furthermore, GLMs do require that the response variable be transformed to normality. Therefore, the data are not forced into unnatural scales, and can allow for non-linearity and non-constant variance in the data. GLMs are based on an assumed relationship (or, the “link” function) between the mean of the response variable and the mean of the linear combination of the potential predictors. Data may be assumed to be from several families of probability distributions, which often fit the non-normal error structures of biological data better than normal distributions.

Using a set of trained GLM prediction models, we identify a list of 160 candidate imprinted genes and experimentally test 10 candidates for evidence of genomic imprinting. Of the 10 genes tested, 2 genes show maternal allele-specific specific expression in the placenta, which is a 60-fold improvement over random chance. Based on this success, we use a microarray based approach to identify a list of candidate imprinted genes, experimentally test an additional 32 genes, and identify 5 more maternally expressed placental genes. Furthermore, the sequence data collected reveals that several histone modifications co-vary significantly with imprinted genes, and that histone modifications are also the strongest predictors of imprinted status. Two recent papers have reported associations between imprinted genes and histone modifications. A paper by Wen *et. al*. looked at two different overlapping chromatin marks, DNA methylation and H3K4me2, and found that regions containing these two overlapping marks showed an approximately 5-fold enrichment for imprinted genes (Wen 2008). While we did not examine levels of DNA methylation or H3K4me2, Mikkelsen *et al.* used the F1 progeny from reciprocal crosses between polymorphic mouse strains in conjunction with ChIP-seq technology to look at enrichment of H3K4me3 and H3K9me3 in imprinted gene regions (Mikkelsen 2007). Of the top 20 regions enriched for both H3K4me3 and H3K9me3, 13 of these mapped to ICRs or imprinted gene promoters. Conversely, of approximately 20 known and putative imprinted loci that contain ICRs, 17 contain these two overlapping chromatin marks. However, in our analysis, H3K4me3 either does not co-vary with imprinted status, or does so in a negative direction, depending on the regions examined. One explanation for this discrepancy is that, while we looked a set intervals spanning 100kb to either side of and within all known genes, Mikkelsen *et. al*.

focused specifically on ICRs. In addition, we did not consider whether the histone modifications overlapped. Rather, we looked at the effect of each histone modification individually.

Although a variety of sequence features were included in our analysis, histone modifications were consistently among the top features examined. For example, H3K9me3, H3K27me3 and H4K20me3 show a strong correlation with imprinted status in a positive, a positive, and a negative direction, respectively. Similarly, the features included in the greatest number of prediction models were mainly histone modifications, although GC percentage and miRNA clusters were included in a large number of prediction models as well. Likewise, the features that were identified as being highly significant in the greatest number of prediction models were the histone modifications H3K27me3 and H3K36me3.

Similarly, although a variety of sequence features were included in our analysis, histone modifications were consistently ranked as important for prediction of imprinted status. These analyses provide insights into regulatory mechanisms controlling imprinting. Our computational studies showed H3K9me3, H3K27me3 and H4K20me3 were the best positive correlates with imprinting. This is consistent with our experimental work demonstrating H3K9me3 and H3K27me3 respectively enforce the methylated and unmethylated states on the parental alleles that are essential to *Rasgrf1* imprinting (Lindroth 2008). It is also consistent with work from others showing that H3K9me3 and H4K20me3 are needed for proper DNA methylation of the maternal allele of *Snrpn* and are also found on the maternal allele of *Rasgrf1*

(Wu 2006; Lindroth 2008; Delaval 2007). H3K36me3 was the best negative correlate with imprinting. This mark is associated with transcriptional repression in yeast and is largely absent from the *Igf2r* promoter in mice (Mikkelsen 2007; Carrozza 2005; Joshi 2005; Keogh 2005). This is significant, as imprinting of *Igf2r* is dependent on transcription of the 103kbp noncoding antisense *Air* transcript across the *Igf2r* promoter (Sleutels 2002). It is likely that the imprinting mechanisms operating at these loci are widely followed at many other imprinted loci that share these four commonly held epigenetic marks.

It is worth mentioning that these histone modification data are not allele specific. Until these data are available, it is unclear whether the histone modifications important for imprinted prediction occur in combination, either on both alleles or on a single allele, or whether one modification marks the active allele and the other marks the repressed allele. It is also important to note that our analysis may be biased towards the identification of placenta imprinted genes, as at least one placenta imprinted gene system (*Kcnq1/Kcnq1ot1*) depends heavily on the presence of histone modifications, which make up the majority of features tested in these models, for imprinted expression (Lewis 2003). Likewise, G9a mutants deficient in both H3K9me2 and H3K9me3 demonstrate placenta-specific imprinting defects, while the embryo proper maintains proper imprinting (Wagschal 2008).

As only 100 out of approximately 25,000 mouse genes are currently known to be imprinted, we would need to test 250 genes to identify a single novel imprinted gene by random chance. Our approach correctly identified a

total of 7 novel imprinted genes out of 42 genes tested in placenta, providing a 40-fold improvement over random chance and indicating that this is a valid method for the identification of novel imprinted genes. This is the first study to link the importance of histone modification data with the successful prediction of novel imprinted genes. As additional data become available on additional epigenetic marks, these methods can compliment transcriptome sequencing and provide insights into the mechanisms controlling genomic imprinting (Wang 2008; Babak 2008).

APPENDIX

The following chapters represent projects that I initiated in parallel. One collaborative project has been published and the rest I did not pursue, as other work described in the body of this thesis proved more promising. However, some of these chapters provide reagents for others to use in the future, some establish new methods and protocols for the lab, and one may be ready for publication with minimal additional work.

III.1 ANTAGONISM BETWEEN DNA AND H3K27 METHYLATION AT THE IMPRINTED *RASGRF1* LOCUS[2]

This work was published in PLoS Genetics in August of 2008. I was involved in the bisulfite sequencing of the DNA regions referred to as D5, D7, and D8. I contributed to the graphs shown in Figure III.1.2 (B, C, and D). In addition, I carried out the chi-square analysis of H3K27me3 and CTCF localization in imprinted versus non-imprinted genes shown in Figure III.1.S2.

III.1.1 Abstract

At the imprinted *Rasgrf1* locus in mouse, a cis-acting sequence controls DNA methylation at a differentially methylated domain (DMD). While characterizing epigenetic marks over the DMD, we observed that DNA and

[2] Lindroth A.M., Park Y.J., McLean C.M., Dokshin G.A., Persson J.M., Herman H., Pasini D., Miró X., Donohoe M.E., Lee J.T., Helin K., Soloway P.D. (2008) Antagonism between DNA and H3K27 methylation at the imprinted *Rasgrf1* locus. *PLoS Genet*. 4, e1000145.

H3K27 trimethylation are mutually exclusive, with DNA and H3K27 methylation limited to the paternal and maternal sequences, respectively. The mutual exclusion arises because one mark prevents placement of the other. We demonstrated this in five ways: using 5-azacytidine treatments and mutations at the endogenous locus that disrupt DNA methylation; using a transgenic model in which the maternal DMD inappropriately acquired DNA methylation; and by analyzing materials from cells and embryos lacking SUZ12 and YY1. SUZ12 is part of the PRC2 complex, which is needed for placing H3K27me3, and YY1 recruits PRC2 to sites of action. Results from each experimental system consistently demonstrated antagonism between H3K27me3 and DNA methylation. When DNA methylation was lost, H3K27me3 encroached into sites where it had not been before; inappropriate acquisition of DNA methylation excluded normal placement of H3K27me3, and loss of factors needed for H3K27 methylation enabled DNA methylation to appear where it had been excluded. These data reveal the previously unknown antagonism between H3K27 and DNA methylation and identify a means by which epigenetic states may change during disease and development.

III.1.2 Author summary

Methylation of DNA and histones exert profound and inherited effects on gene expression. These occur without changes to the underlying DNA sequence and are considered epigenetic effects. Disrupting epigenetic states can cause developmental abnormalities and cancer. Very little is known about how locations in the mammalian genome are chosen to receive these chemical modifications, or how their placement is regulated. We have

identified a DNA sequence that acts as a methylation programmer at the *Rasgrf1* locus in mice. It is required for methylation of nearby DNA sequences and can also influence the levels of local histone methylation. The methylation programmer has different effects on paternally and maternally derived chromosomes, directing DNA methylation on the paternal allele and histone H3 lysine 27 trimethylation on the maternal allele. These two methylation marks are not only mutually exclusive; they are also mutually antagonizing, whereby one blocks the placement of the other. Manipulations that cause aberrant changes in the levels of one of these marks had the opposite effect on the other mark. These observations identify novel mechanisms that specify epigenetic states in vivo and provide a framework for understanding how pathological epigenetic changes can arise, including those emerging at tumor suppressors during carcinogenesis.

III.1.3 Introduction

In mammals, imprinted loci are expressed from only one allele. Accompanying and controlling monoallelic expression are allele-specific epigenetic modifications influenced by an imprinting control region (ICR). Within this region, there is a differentially methylated domain (DMD) that is subject to acquisition of epigenetic modifications, typically DNA methylation and histone modifications. These modifications are placed in a parent-of-origin specific manner and impose an epigenetic state that dictates allele-specific gene expression at imprinted loci (Edwards 2007).

Previously, we characterized the mechanisms by which the ICR controls allele-specific methylation and expression at the imprinted *Rasgrf1*

locus. The ICR, located 30 kbp upstream of the transcriptional start site, is a binary switch consisting of a repeated element and the DMD. The repeated element functions as a methylation programmer, that is necessary for the establishment and maintenance of DNA methylation at the DMD on the paternal allele and sufficient for establishing gametic imprints in both germlines (Yoon 2002; Holmes 2006; YJP, HH, AML, Ying Gao and PDS, in preparation). The DMD is a methylation sensitive enhancer blocker that binds CTCF on the unmethylated maternal allele and limits enhancer to promoter interactions, silencing the maternal allele (Yoon 2005). DNA methylation that is directed to the paternal DMD by the repeats prevents CTCF binding, allowing expression of the paternal allele. The repeats constitute the first identified, and one of only a few known, naturally occurring DNA methylation programmers in mammals (Shemer 2000; Bastepe 2003; Linglart 2005; Perk 2002).

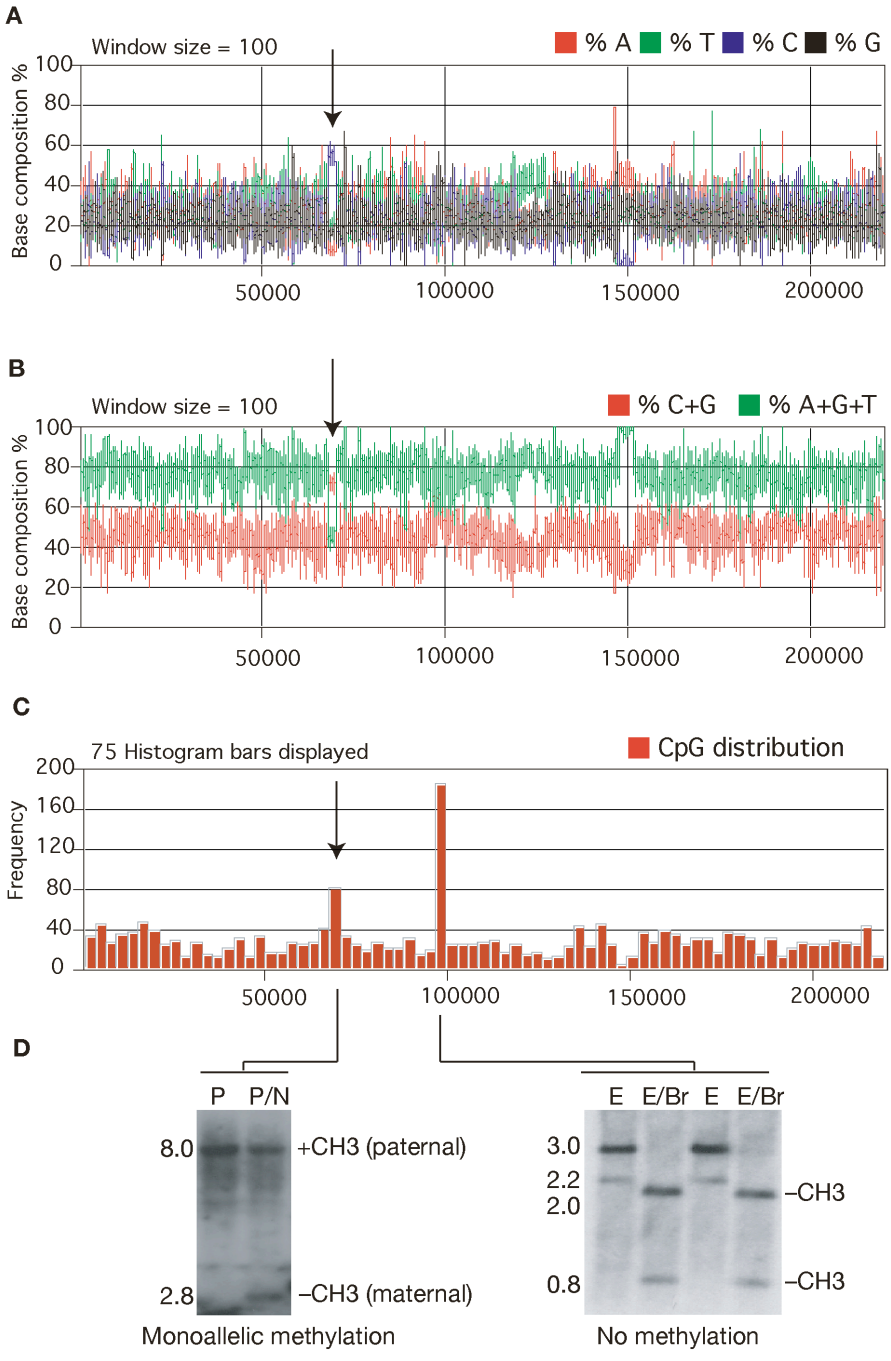
Epigenetic analysis of *Rasgrf1* done by others examined DNA methylation across an expanded region centered on the ICR (Kobayahshi 2006; Hisato Kobayashi and Hiroyuki Sasaki unpublished data) and histone modifications at the ICR (Delaval 2007). The DNA methylation data suggested that a broader DMD exists in somatic tissue and in the male germline than was previously appreciated (Kobayashi 2006). The histone methylation data indicated that several allele-specific histone modifications accompany the DNA methylation differences, including H3K27me3 and H4K20me3 on the maternal allele and H3K9me3 on the paternal allele (Delaval 2007).

The *Rasgrf1* locus presents some unusual paradoxes: The paternal

allele is active yet it carries DNA methylation and other repressive marks, whereas the maternal allele is silent and lacks DNA methylation but carries other repressive marks. It is unclear if and how the primary DNA sequence controls each of these parent-specific marks. We have identified the DNA sequences that are necessary and sufficient for programming the establishment and maintenance of DNA methylation on the paternal allele, however, nothing is known about the cis-acting DNA sequences that control placement of repressive histone modifications in this region, or whether there is any coordination between the histone and DNA methylation modifications. In many organisms, distinct epigenetic marks coordinately determine the transcriptional status of genes. For instance, recruitment of DNA methylation can depend upon pre-established histone H3 methylation at lysine 9 (Tamaru 2003; Johnson 2002; Lehnertz 2003); histone modifications can be lost when DNA methylation is impaired (Espada 2004); and some histone modifications become redistributed in histone methyltransferase mutants (Peters 2003).

Here we report the analysis of a 12 kbp region at *Rasgrf1* for locations bearing histone modifications and DNA methylation. Our data reveal the mutual exclusion of the repressive H3K27 methylation and DNA methylation modifications. Furthermore, by experimentally manipulating the levels of DNA and H3K27 methylation possible at the locus, we demonstrate that these two marks are mutually antagonistic, whereby the placement of one mark prevents the placement of the other, and removal of one mark allows the encroachment of the other. Additionally, we found that the tandem repeat sequences, which are necessary and sufficient for programming DNA methylation marks, are also important for directing H3K27 and H3K9 modifications to the proximal

Figure III.1.S1. CpG dinucleotides and methylated DNA centered at the DMD. A,B,C. The distribution of A, T, C and G over the 220 kb cluster show that there is a predominant accumulation of cytosines over the DMD and repeat region. The DMD repeat region has large amount of C and G together. There are two CpG islands in the region: one, which is the DMD and the other, which is in the promoter region of *Rasgrf1* (bottom panel). D. Southern blot analysis of the two CpG islands using methylation sensitive restriction enzymes and tail DNA show that there is monoallelic methylation at the DMD (left panel) but no methylation of the promoter region CpG island (right panel). P (PstNI), N (NotI, methylation sensitive), E (EcoRI), Br (BsrBI, methylation sensitive). Bands diagnostic for the methylated (+) and unmethylated (–) states are indicated.


A
Window size = 100
% A
% T
% C
% G
Base composition %
100
80
60
40
20
0
50000
100000
150000
200000
B
Window size = 100
% C+G
% A+G+T
Base composition %
C
75 Histogram bars displayed
CpG distribution
Frequency
200
160
120
80
40
0
D
P
P/N
8.0
+CH3 (paternal)
2.8
–CH3 (maternal)
Monoallelic methylation
E
E/Br
E
E/Br
3.0
2.2
2.0
–CH3
0.8
–CH3
No methylation

DMD and that H3K9 methylation is needed for optimum establishment of DNA methylation on the paternal allele.

III.1.4 Results

III.1.4.1 The *Rasgrf1* repeats program DNA methylation only at the 400nt DMD in the male germline

There are two regions rich in C and G residues and CpG dinucleotides over a 200 kbp interval at the *Rasgrf1* locus. One CpG cluster is in the ICR and the other in the promoter region of *Rasgrf1* (Figure III.1.S1A, III.1.S1B, III.1.S1C). By analyzing the DNA methylation pattern of these two CpG clusters in somatic DNA using methylation sensitive restriction enzymes, we found that only the DMD CpG cluster is methylated while the one in the promoter is not (Figure III.1.S1D). When Kobayashi et al. performed a comprehensive analysis analysis of allele-specific DNA methylation at the *Rasgrf1* ICR in embryonic day 12.5 DNA and in the male germline, they observed that the somatic and germline DMD was larger than had been previously appreciated (Kobayashi 2006; Hisato Kobayashi and Hiroyuki Sasaki unpublished data). We expanded upon this by characterizing the distribution of both histone modifications and DNA methylation over a 12 kbp region centered on the DMD within the ICR, and also by evaluating the influence of the tandem repeats within the ICR on these epigenetic marks (Figure III.1.1).

We performed bisulfite sequencing to characterize DNA methylation in 86 CpGs present in eight segments (labeled D1 through D8 in Figure III.1.1)

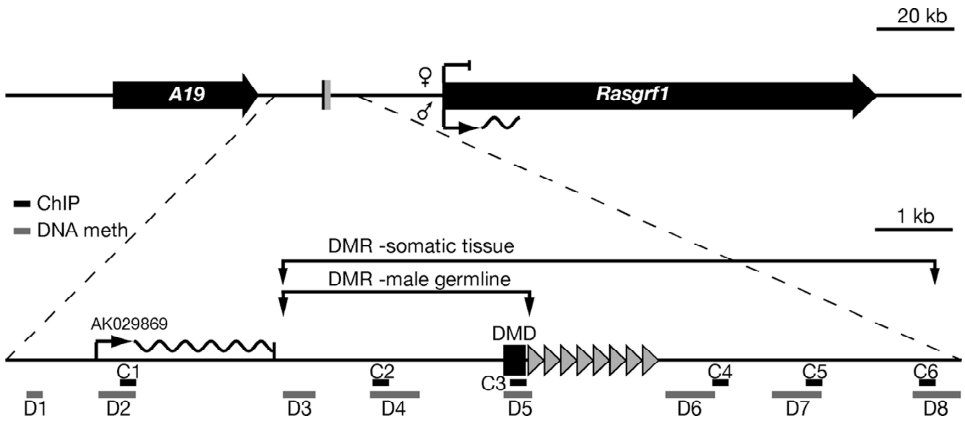


Figure III.1.1. Schematic view of the imprinted *Rasgrf1* locus. A 220 kbp region (top) contains the paternally-expressed A19 and *Rasgrf1* transcripts (black, rightward pointing arrows) and the ICR (black rectangle). Detailed view of 12 kbp centered on the ICR (bottom) includes the originally defined 400 nt core differentially methylated domain (DMD) and tandem repeats (triangles), which constitute the DNA methylation programmer. Amplicons for ChIP (C1–C6) and DNA bisulfite sequencing (D1–D8) are indicated. The sites of the germline and somatic DMR as described by Kobayashi et al. are shown (Kobayashi 2006; Hisato Kobayashi and Hiroyuki Sasaki, unpublished).

Figure III.1.2. Distribution of DNA methylation over 12 kbp spanning the ICR. (A) Alleles analyzed. The functionally wildtype (WT-flox) allele has a loxP sites (black rectangles) flanking the tandem repeat element (triangles) in the ICR and behaves like an unmodified wildtype (+) allele (Holmes 2006). *Cre* recombinase was used to delete the repeat element (RepΔ). Primers P1–P2 amplify the C3 region from + or WT-flox alleles; P3 and P4 amplify the RepΔ allele. (B) CpG methylation in spermatozoa DNA isolated from wildtype (+/+) mice or mice homozygous for the repeat element deletion (RepΔRepΔ). (C) Maternal allele CpG methylation in neonatal brain of mice that harbor two wildtype (+/+) alleles, a paternal repeat element deletion (+/RepΔ) or maternal repeat element deletion (RepΔ/+). Maternal allele is shown first. (D) Paternal allele CpG methylation of DNAs from (C). The fraction of CpGs methylated in each segment (D1 through D8) is reported, calculated as the fraction of CpGs methylated in all clones we sequenced from bisulfite treated DNAs. A total of 500 templates were sequenced. Table III.1.S1 reports the number of sequences for the two parental alleles in each of the regions assayed (D1–D8) in the various somatic and germline DNAs.

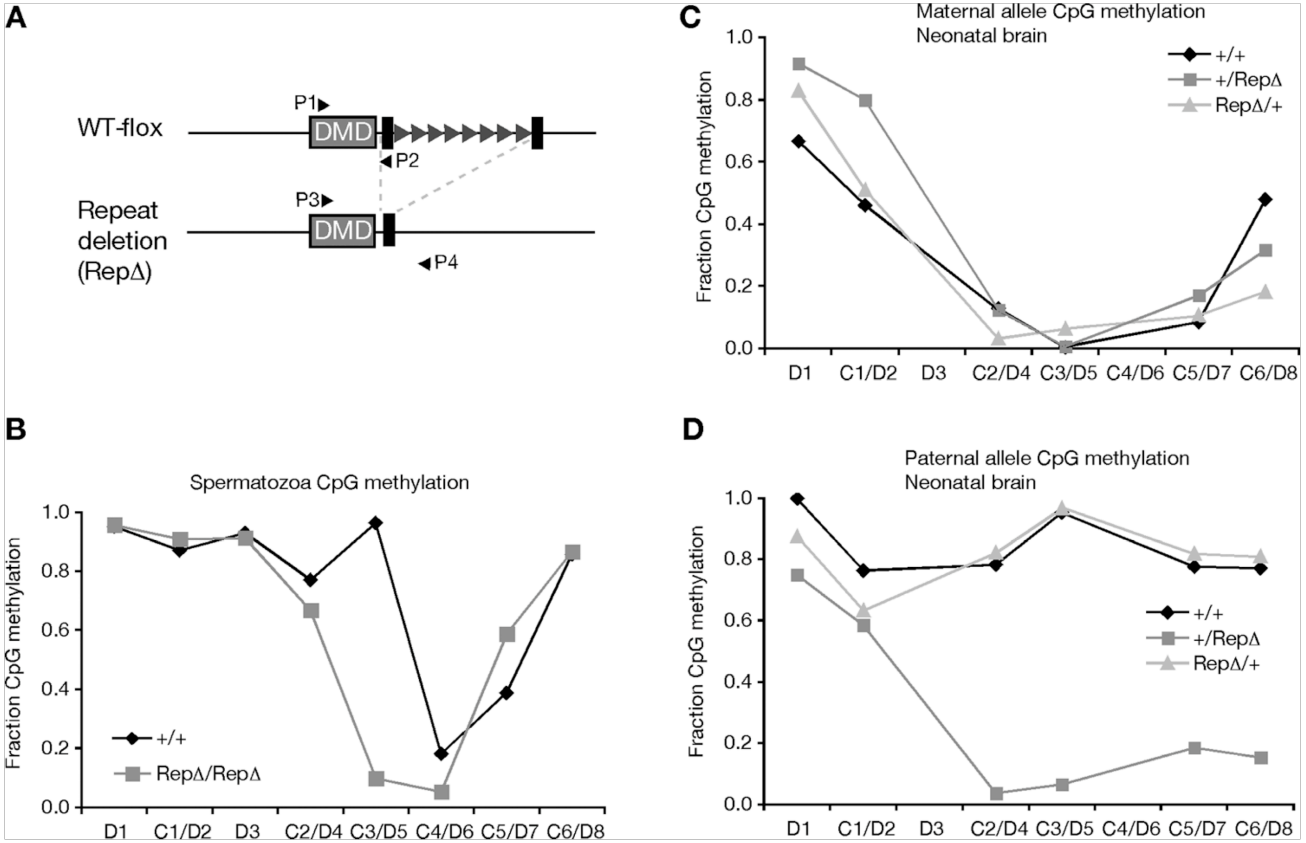

A
WT-flox
P1
DMD
P2
Repeat
deletion
(RepΔ)
P3
DMD
P4
B
Spermatozoa CpG methylation
Fraction CpG methylation
1.0
0.8
0.6
0.4
0.2
0.0
+/+
RepΔ/RepΔ
D1
C1/D2
D3
C2/D4
C3/D5
C4/D6
C5/D7
C6/D8
C
Maternal allele CpG methylation
Neonatal brain
Fraction CpG methylation
+/+
+/RepΔ
RepΔ/+
D
Paternal allele CpG methylation
Neonatal brain
Fraction CpG methylation
+/+
+/RepΔ
RepΔ/+

Table III.1.S1. Clones sequenced for analysis in Figure 2. DNAs from neonatal brains, taken from mice with the three indicated genotypes, and from sperm with the two indicated genotypes, were subjected to bisulfite PCR and the PCR products were cloned and sequenced. Primers used to amplify regions D1 through D8 are listed in Table S1. This table reports the number of clones sequenced that correspond to the maternal and paternal alleles for brain DNA. Assignment of individual clones to the maternal or paternal allele required the use of polymorphisms between PWK and 129S4Jae parents of F1 DNAs, as described in Supporting Methods. Note that no polymorphisms were present in D3 and D6 so allele specific methylation was not determined (nd) in neonatal brain DNA from F1 mice in those regions.

		D1	D2	D3	D4	D5	D6	D7	D8	TOTALS
NEONATAL BRAIN										
+/+	Maternal	3	20	nd	16	11	nd	32	17	99
	Paternal	3	13	nd	16	9	nd	13	17	71
+/RepΔ	Maternal	3	2	nd	9	6	nd	8	17	45
	Paternal	4	18	nd	9	8	nd	7	9	55
RepΔ/+	Maternal	2	13	nd	7	10	nd	18	18	68
	Paternal	4	6	nd	10	7	nd	5	7	39
SPERM										
+/+		7	7	9	7	10	8	5	7	52
RepΔ/RepΔ		8	7	10	16	9	11	8	13	71
TOTALS		34	86	19	90	70	19	96	105	500

containing 4,118 bp from the 12,020 bp interval. Our analysis of methylation in the soma used DNA from neonatal brain and our analysis in the male germline used DNA isolated from sperm. Somatic DNAs were from F1 progeny of 129S4Jae and PWK strains. Polymorphisms between these strains allowed us to determine which bisulfite sequences were from the maternal and paternal alleles in the soma. The 129S4Jae-derived allele was either wildtype or lacked the *Rasgrf1* tandem repeats constituting the DNA methylation programmer (Figure III.1.2A). Sperm DNAs were from mice homozygous for wildtype or tandem repeat-deficient alleles of *Rasgrf1*. Our characterization of the somatic methylation states from animals carrying the wildtype 129S4Jae allele was in strong agreement with the results of Kobayashi, even though sources of somatic DNAs differed: Kobayashi used midgestation embryos. In neonatal brain DNA, we detected paternal allele-specific DNA methylation, which covers at least the 7.6 kbp interval between segments D4 through D7 and includes the ICR. We also found a region of methylation on both alleles over a 1.4 kbp interval upstream of the ICR containing segments D1 and D2. None of the somatic DNA methylation patterns changed on either the paternal or maternal alleles in mice harboring a deletion of the tandem repeats on the maternal allele (Figure III.1.2C, III.1.2D). In contrast, all paternal allele-specific DNA methylation we detected in regions D4 to D8 was lost from somatic DNA when the tandem repeats were absent from the paternal allele (Figure III.1.2D). This indicates that the range of action of the *Rasgrf1* DNA methylation programmer within the tandem repeats is not confined to the narrowly defined 400nt DMD previously studied, but its reach spans at least 7 kbp in somatic tissue.

Imprinted DNA methylation patterns that are established in the germlines are typically maintained and can even spread during somatic development. To determine the extent of the methylation in sperm DNA and the range of action of the DNA methylation programmer in the male germline, we performed bisulfite analysis on sperm DNA from mice carrying an intact repeat element and also from mice carrying a deletion of the repeats. In mice with the intact repeats, we found that *Rasgrf1* methylation in sperm DNA was present not only the originally defined 400bp DMD, but it extended an additional 1.6 kbp upstream, in agreement with results from Hisato Kobayashi and Hiroyuki Sasaki (unpublished). However, in mice bearing a deletion of the repeats that constitute the *Rasgrf1* DNA methylation programmer, only the DNA methylation at the originally defined DMD was lost. The DNA methylation on the additional 1.6 kbp was unaffected, indicating that the range of action of the *Rasgrf1* DNA methylation programmer in the tandem repeats is limited to the 400 bp proximal DMD in the male germline (Figure III.1.2B). Because loss of DNA methylation on that narrowly defined sequence was sufficient to disrupt imprinted expression of *Rasgrf1* (Yoon 2002), we infer that this differentially methylated portion of the locus is essential for its imprinting and we refer to it as the core DMD.

III.1.4.2 Allele-specific histone modifications are present at the DMD and depend on the DNA methylation programmer

We next characterized histone methylation status across the same 12 kbp interval over which the DNA methylation was characterized. Specifically, we sought to determine where histone modifications were distributed, if any modifications were allele-specific, if their placement required the same DNA

methylation programmer that imprinted DNA methylation requires, and if there is any coordination between modification states on histones and DNA. We limited our analysis to di-, and tri-methylation of histone H3 at lysine 9 and 27 because they are associated with gene silencing and DNA methylation, which are observed at the maternal and paternal alleles respectively. For this analysis, we performed chromatin immunoprecipitation (ChIP) using mouse embryonic fibroblasts (MEFs) and antibodies specific to H3K9me2, H3K9me3, H3K27me2, and H3K27me3. Our initial tests were controls to verify that the antibodies detected histone modifications with proper specificity (Figure III.1.S2). For these tests, we amplified immunopreciptates using primers from Charlie, Actin and Hoxa9. H3K9me2 and H3K27me2 are known to reside at Charlie (Martens 2005), H3K9me3 at Actin, and H3K27me3 at Hoxa9 (Schlesinger 2007). The expected PCR products were observed for each immunoprecipitation, indicating the antibodies were indeed specific. In addition, PCRs done using DMD primers detected only H3K9me3 and H3K27me3 at the DMD; therefore, subsequent ChIP studies primarily used antibodies recognizing these marks (Figure III.1.S2).

We then extended our H3K27me3 and H3K9me3 analysis to six segments (labeled C1 through C6 in Figure III.1.1) that included 10,451 bp surrounding the core DMD and methylation programmer using two separate immunopreciptates from wildtype MEFs, and MEFs carrying a deletion of the DNA methylation controlling repeats (RepΔ). Because we used two immunoprecipitates, these analyses report the general distribution of histone marks in the region rather than providing reliable quantification of their abundance. Our PCRs in regions C1, C2, C4, C5 and C6 did not distinguish

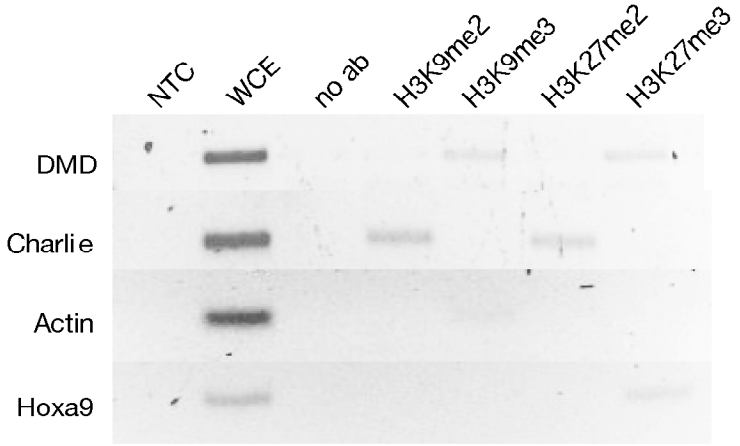


Figure III.1.S2. Specificity controls for antibodies used in ChIP. Representative gel analysis of ChIP results indicating the specificity of the antibodies for the histone modification analysis in this study. Antibodies specific to H3K9me2, H3K9me3, H3K27me2, and H3K27me3 show enrichment for H3K9me3 and H3K27me3 at the DMD. Positive control PCRs for H3K9me2 and H3K27me2 (Charlie), H3K9me3 (Actin), and H3K27me3 (Hoxa9) are included as well as a test for the *Rasgrf1* DMD. NTC, no template control; WCE, whole cell extract not immunoprecipitated; no ab, mock precipitations done without antibody.

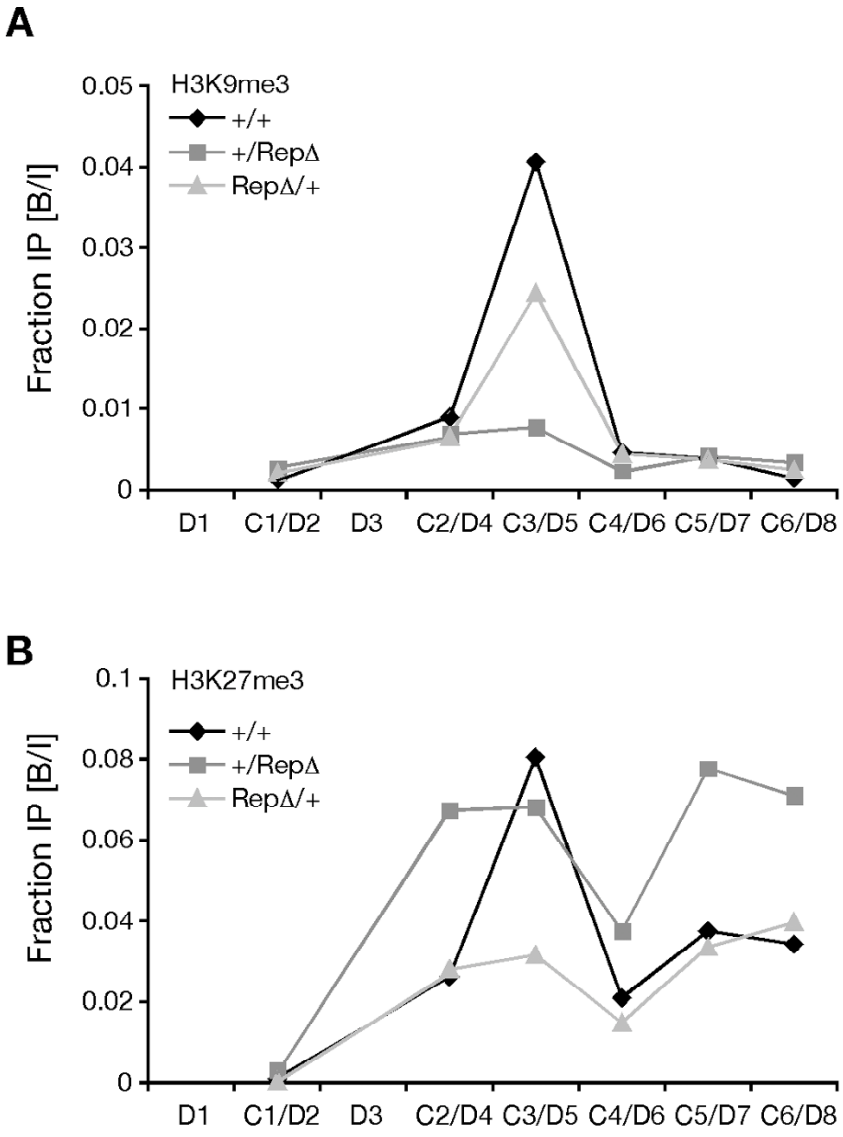
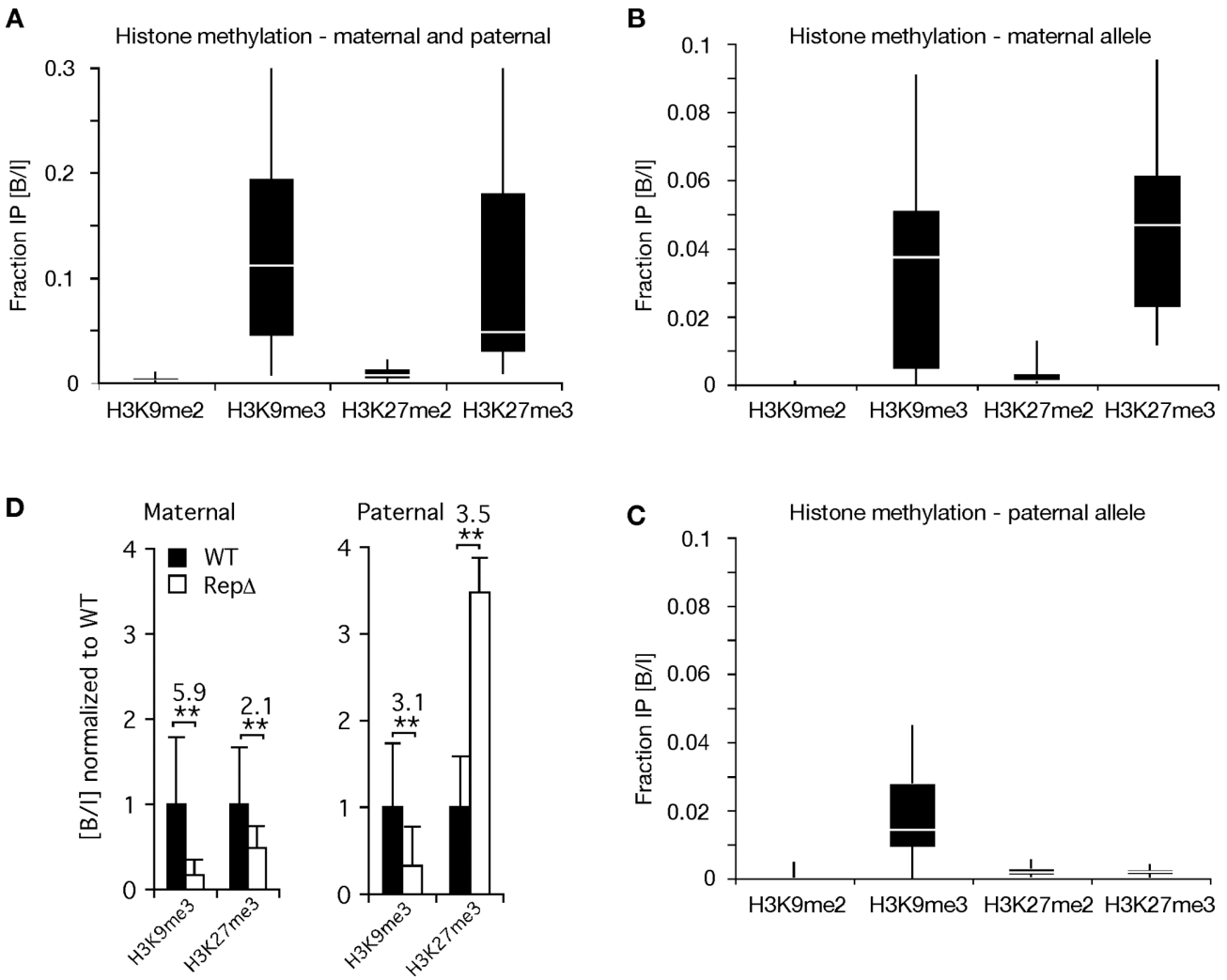


Figure III.1.3. Distribution of methylated histones at *Rasgrf1*. (A) H3K9me3 and (B) H3K27me3 distribution over 12 kbp spanning the ICR in MEFs isolated from wildtype (+/+) mice and mice with a deletion of the tandem repeat element from the paternal (+/RepΔ) or maternal (RepΔ/+) alleles. Data plotted as fraction bound over input, as measured by quantitative real-time PCR (Q-PCR). Note that in +/RepΔ MEFs, which had lost DNA methylation, the H3K27me3 mark encroached into sites where it was originally absent in +/+ MEFs.

the parental alleles and our PCR of the DMD at C3 used wildtype allele-specific primers. The ChIP analysis of wildtype MEFs indicated that both H3K9me3 and H3K27me3 were most abundant at the core DMD at region C3 with some H3K27me3 signal extending downstream of the tandem repeats (Figure III.1.3). Analysis of MEFs carrying a deletion of the DNA methylation controlling repeats suggested that the repeats could influence the distribution of histone modifications at the DMD and elsewhere in the region.

To provide statistically significant measures of methylated H3K9 and H3K27 at the DMD and to assess if any modifications were allele-specific, we analyzed a total of six to twelve independent immunoprecipitations by quantitative PCR using primers spanning the DMD at region C3 (Figure III.1.2A). Our data confirmed that the DMD is enriched for trimethylated lysines but lacks dimethylated ones (Figure III.1.4A). To determine if these histone marks over the DMD were on the maternal or paternal alleles, we repeated the ChIP assays using mice carrying the engineered polymorphisms shown in Figure III.1.2A that enabled us to amplify the wildtype maternal and paternal DMD sequences separately. Results demonstrated that the maternal allele has H3K9me3 and H3K27me3, whereas the paternal DMD has only the H3K9me3 mark (Figure III.1.4B and III.1.4C). This is in partial agreement with other data describing H3K27me3 as being maternal allele specific and H3K9me3 as being paternal allele specific at *Rasgrf1* (Delaval 2007). H3K9me3 that we detect on the two alleles may be placed by different mechanisms. Our data correlate well with previous findings that DNA methylation can be coregulated with H3K9me3 (Tamaru 2003; Lehnertz 2003; Malagnac 2002; Jackson 2002), but generally not with H3K27me3 (Zhang

Figure III.1.4. Allele specific histone modifications on the *Rasgrf1* core DMD and their sensitivity to methylation programming repeats. (A–D) Quantitative PCR analysis of immunoprecipitates and unprecipitated input materials was used to calculate the fraction of input material precipitated by each antibody (reported as Bound/Input, B/I). Box plots report distribution of B/I values from 6–12 individual precipitations, each analyzed in triplicate. (A) Analysis of the DMD using wildtype cells, which does not distinguish the alleles, shows the DMD is enriched for H3K9me3 and H3K27me3 (one-way ANOVA and multiple comparison: Hsu's MCB, a = 0.05). (B) Maternal allele specific analysis of the wildtype DMD using wildtype allele-specific primers and immunoprecipitates from heterozygous +/WT-flox or +/RepΔ cells. Modified alleles are shown in Figure 2A. (C) Paternal allele specific analysis of the wildtype DMD performed as in B but using WT-flox /+ or RepΔ/+ cells. In B and C, results did not depend on which heterozygote was used indicating that no interactions between the alleles affected the histone modifications analyzed. H3K9me3 is present on both alleles, while H3K27me3 is present only on the maternal allele ($p<0.01$). (D) Sensitivity of H3K9me3 and H3K27me3 methylation to the tandem repeats. Graph represents B/I values from ChIP analysis of the RepΔ allele, obtained using primers P3/P4 in Figure 2A (white bars), normalized to measurements from the wildtype (WT) allele, obtained using primers P1/P2 in Figure 2A (black bars). Analyses used RepΔ/+ and +/RepΔ cells to assess the importance of the maternal and paternal repeats respectively. Deletion of the maternal repeats caused significant decreases in H3K9me3 and H3K27me3 on the maternal allele (**, $p<0.01$). Deletion of the paternal repeats caused a significant decrease in paternal H3K9me3 (**, $p<0.01$) and an increase in paternal H3K27me3 (**, $p<0.01$). Fold increase or decrease is indicated above each pair of bars. Deletions of the repeat on one allele did not affect the histone states on the homologous allele (not shown). P values determined by a mixed model with log-transformed B/I values, a fixed independent variable for allele (wildtype or repeat deletion) and a random variable IP.


A
Histone methylation - maternal and paternal
Fraction IP [B/I]
0.3
0.2
0.1
0
H3K9me2
H3K9me3
H3K27me2
H3K27me3
B
Histone methylation - maternal allele
Fraction IP [B/I]
0.1
0.08
0.06
0.04
0.02
0
H3K9me2
H3K9me3
H3K27me2
H3K27me3
D
Maternal
Paternal
WT
RepΔ
[B/I] normalized to WT
4
3
2
1
0
5.9
**
2.1
**
3.1
**
3.5
**
H3K9me3
H3K27me3
H3K9me3
H3K27me3
C
Histone methylation - paternal allele
Fraction IP [B/I]
0.1
0.08
0.06
0.04
0.02
0
H3K9me2
H3K9me3
H3K27me2
H3K27me3

2007; Mathieu 2005).

Because the tandem repeats act as a DNA methylation programmer, playing an essential role both in establishment and maintenance of DNA methylation at the DMD (Yoon 2002; Holmes 2006; YJP, HH, AML, Ying Gao and PDS, in preparation), we wanted to determine if they also influence placement of methylated histone marks at the DMD. We did this by repeating the ChIP analysis using MEFs carrying a deletion of the repeats (RepΔ) and amplifying the wildtype allele and the mutated allele separately. Our analysis showed that the repeat element indeed has a significant influence of histone modification status at the DMD, in addition to controlling its DNA methylation (Figure III.1.4D): When the repeats were absent from the maternal allele, the levels of maternal allele-specific H3K27me3 and H3K9me3 were respectively 1/2 and 1/6th the levels seen when the repeats were present. Similarly, when the repeats were absent from the paternal allele, the level of paternal allele-specific H3K9me3 was 1/3rd that seen when the methylation programmer was absent. Interestingly, deletion of the repeats from the paternal allele led to a three-fold increase in the accumulation of H3K27me3 on the paternal allele. This is consistent with our locus wide ChIP analysis spanning intervals C1 to C6, which suggested H3K27me3 can encroach into areas where it is normally absent, both 5′ and 3′ of the DMD, when the paternal repeats are deleted (see sites C2, C5, C6 in Figure III.1.3B). These observations provide evidence that DNA methylation and H3K27me3 are mutually exclusive epigenetic marks at *Rasgrf1*. When we superimposed the DNA and H3K27 methylation data for wildtype animals and animals carrying a deletion of the paternal methylation programmer from Figures III.1.2 and III.1.3, the mutual exclusion of

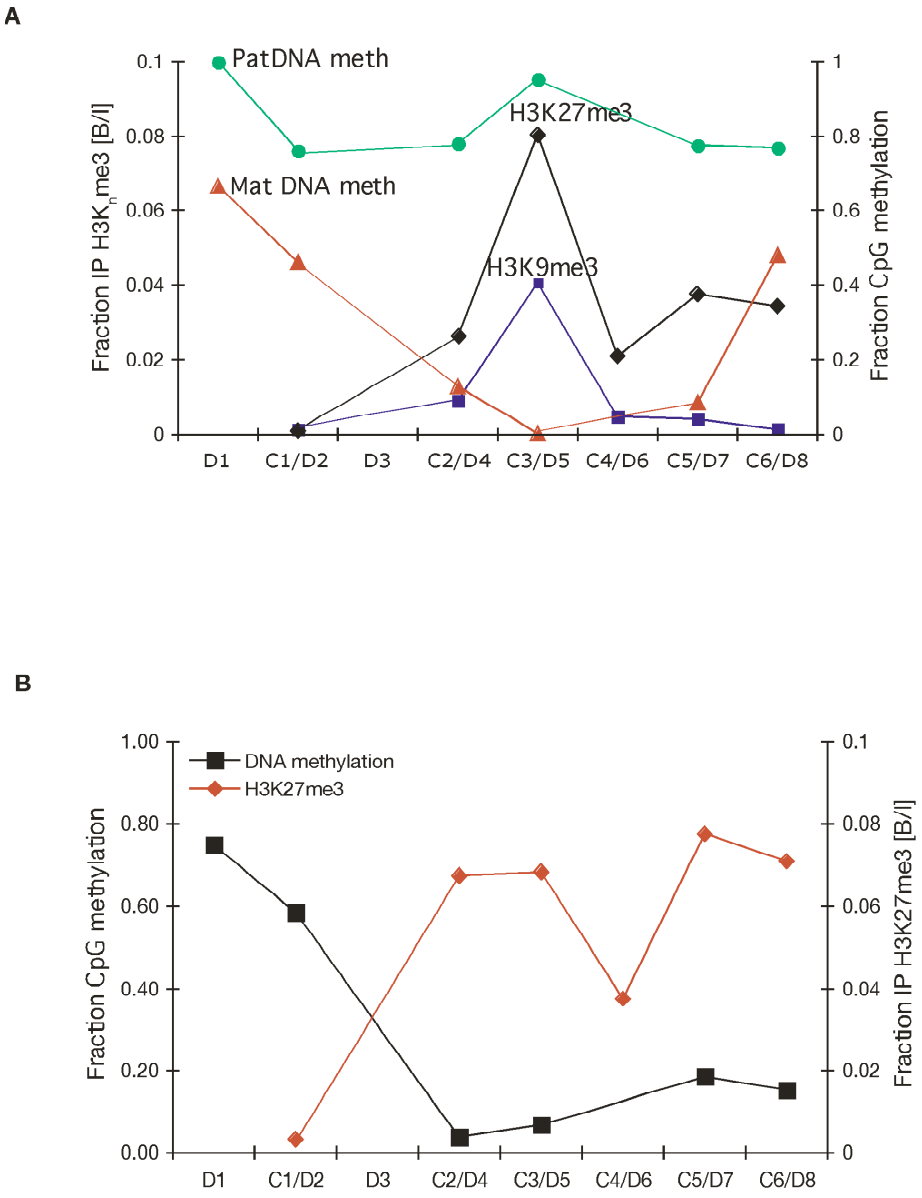


Figure III.1.S3. Mutual exclusion of H3K27 and DNA methylation. H3K27 and DNA methylation data from figures 2 and 3 were redrawn to highlight the mutual exclusion of H3K27me3 and DNA methylation. (A) Modifications present at *Rasgrf1* in wildtype MEFs show that paternal DNA methylation (green) is largely even over the region, while maternal DNA methylation (red) is absent over the DMD but present upstream and downstream. Strikingly, H3K9me3 and H3K27me3 are perfectly confined to the DMD. (B) Modifications in the paternal allele in +/RepΔ mice. DNA methylation (black) is lost from the DMD and downstream, allowing encroachment of H3K27me3 into these regions (red).

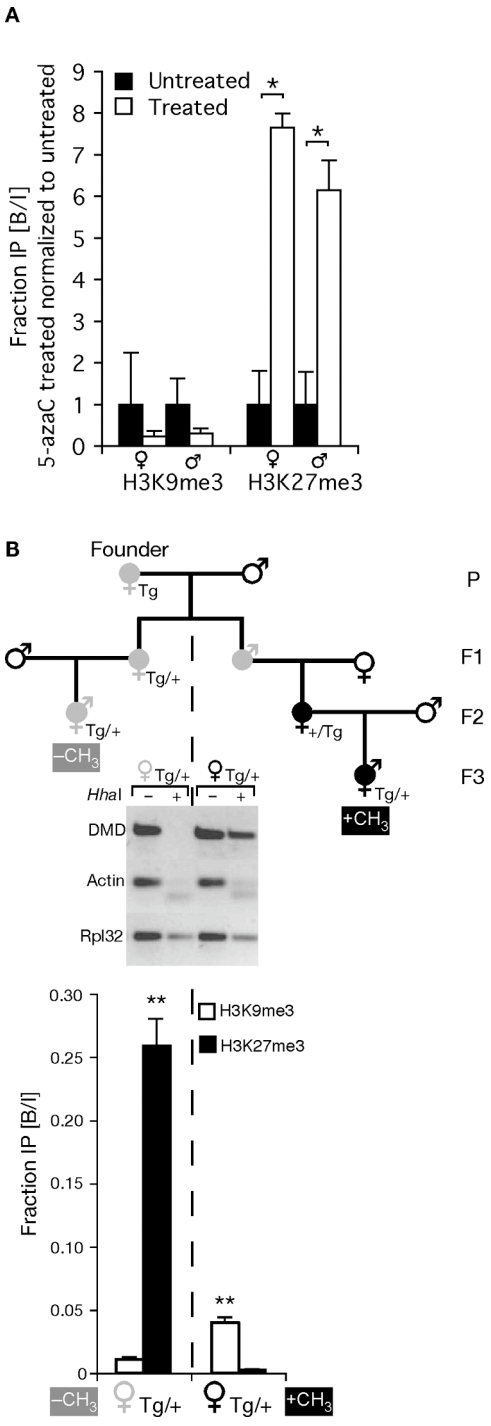
H3K27me3 and DNA methylation over the core DMD was apparent (shown separately in Figure III.1.S3A and III.1.S3B for clarity).

III.1.4.3 Antagonism between H3K27 and DNA methylation

Mutual exclusion of H3K27me3 and DNA methylation can arise by different mechanisms. In one scenario, the two marks may be placed in different compartments of the nucleus and the DNA cannot reside in both places. Alternatively, distinct factors needed for H3K27me3 and DNA methylation may require the same DNA binding site, which cannot be simultaneously occupied by the two sets of factors. A third possibility is that DNA and H3K27 methylation are mutually antagonizing, whereby one inhibits placement of the other. This last possibility is mechanistically different from mere mutual exclusion. If antagonism between these two marks is occurring, then we can predict what happens to one mark if the other is experimentally manipulated.

In order to explore more directly the possible antagonism between DNA methylation and H3K27me3, we repeated our allele-specific ChIP studies using MEFs that had been treated with the DNA methyltransferase inhibitor 5-azacytidine. If DNA methylation can antagonize H3K27 methylation, then we expected that 5-azacytidine treatments should increase the levels of H3K27me3 at the DMD as assayed by ChIP. This is precisely what we observed. 5-azacytidine treatments increased the signals from our ChIP analysis by more than six fold when we assayed H3K27me3 on the two parental alleles (Figure III.1.5A). Although the maternal allele lacks imprinted DNA methylation, there is DNA methylation at sites D1, D2 and D8.

Figure III.1.5. DNA methylation excludes H3K27me3 from the *Rasgrf1* ICR, while loss of DNA methylation lead to acquisition of H3K27me3. (A) Allele specific ChIP analyses using MEFs treated for 24 hours with 0.4 μM of 5-azacytidine (5-azaC), a DNA methyltransferase inhibitor. B/I values from treated cells were normalized to the values from untreated control cells. H3K27me3 was significantly increased on both alleles after 5-azaC treatment (* $p<0.05$). (B) Transgene specific ChIP analyses. Upper panel: Pedigree of transgenic mice from which MEFs were isolated. Transgene positive animals indicated by filled grey or black symbols. A female parental generation (P) transgenic founder and her F1 progeny have an unmethylated transgene (filled gray symbols). Strict maternal inheritance of the transgene preserved its unmethylated state (left portion of pedigree), whereas any intervening transmission by males caused methylation (filled black symbols) that persisted even after subsequent transgene transmission by females (right portion of pedigree). Middle panel: Methylation state at five HhaI sites in the transgene. DNAs were prepared from MEFs grown from the last progeny of both halves of the pedigree shown in A that carried a maternally transmitted transgene. DNAs were undigested (–) or digested with the methylation-sensitive enzyme HhaI (+) prior to PCR amplification using transgene specific primers. As a control for HhaI digestion, DNAs were amplified using primers from Actin that flank an unmethylated HhaI site. As a control for the PCR reaction, DNAs were amplified using primers from Rpl32 that span an interval lacking HhaI sites. Lower panel: ChIP analysis for H3K9me3 and H3K27me3 using MEFs whose maternal transgenic DMD was methylated (black □ symbol, right two bars) or unmethylated (grey □ symbol, left two bars). There is a significant enrichment of H3K27me3 on the unmethylated maternal transgene and H3K9me3 on the methylated maternal transgene (** $p<0.01$). P values determined by a mixed model with log-transformed B/I values, a fixed independent variable for 5-azaC treatment (A) or parental transmission (B) and a random variable IP.


A
Untreated
Treated
Fraction IP [B/I]
5-azaC treated normalized to untreated
H3K9me3
H3K27me3
B
Founder
P
F1
F2
F3
HhaI
DMD
Actin
Rpl32
H3K9me3
H3K27me3
Fraction IP [B/I]

Reductions in methylation at those sites might augment accumulation of H3K27me3 across the entire region. If DNA methylation antagonizes H3K27 methylation, then an additional expectation is that inappropriate placement of DNA methylation on the maternal DMD should exclude accumulation of H3K27me3 marks. To test this, we took advantage of a transgenic system we developed to test if the tandem repeats, which are necessary for programming DNA methylation at *Rasgrf1*, are sufficient to impart imprinted methylation to the DMD at an ectopic location in the genome. Independent transgenic founders harboring three to five ectopic copies of the *Rasgrf1* ICR underwent proper establishment of DNA methylation at the transgenic DMD in the male germline and erasure of that methylation in the female germline, recapitulating the essential features of imprinted methylation establishment seen at the endogenous locus (YJP, HH, AML, Ying Gao and PDS in preparation). We were able to distinguish the transgenic ICR from the endogenous copy because the transgenic repeats were flanked with loxP sites and had the same structure as the WT-flox allele shown in Figure III.1.2A. This allowed us to assay DNA methylation and perform ChIP analysis of the transgene. The transgene was useful for the studies we describe here because the unmethylated state that was established on the transgene after female transmission could not be maintained if there was any history in the pedigree of transmission through a male (Figure III.1.5B). This system of transgenerational epigenetic memory allowed us to generate two different sets of MEFs, both of which were derived by maternal transmission of the transgene from a common founder. For one set of MEFs, the transgene was unmethylated at the transgenic DMD, whereas in the second set, the same transgene was methylated on the DMD (Figure III.1.5B, upper two panels). If

there is antagonism between H3K27me3 and DNA methylation at *Rasgrf1*, then we predicted our MEFs with a methylated transgene should exclude H3K27me3, whereas our MEFs with an unmethylated transgene should allow its placement. This is also precisely what we observed (Figure III.1.5B, lower panel), providing additional independent evidence that DNA methylation antagonizes placement of H3K27me3.

We next wondered if the antagonism between DNA methylation and H3K27me3 might be reciprocal, meaning; H3K27me3 is able to exclude DNA methylation. To test this possibility we analyzed the DNA methylation state of the *Rasgrf1* DMD in ES cells, embryoid bodies or trophoblast outgrowths that lack either of two factors needed for H3K27me3 by the PRC2 complex. PCR2 includes SUZ12, EED and EZH2, the H3K27 methyltransferase. YY1, the mammalian ortholog of the Drosophila PHO protein, is a DNA binding factor that binds to EED and recruits PRC2 to sites of action (Satijn 2001; Srinivasan 2005; Wilkinson 2006). Mice and cells with deficiencies in either SUZ12 or YY1 fail to acquire normal levels of H3K27me3 genome wide (Satijn 2001; Caretti 2004; Pasini 2004), and the deficiency is lethal for mice, but SUZ12-deficient ES cells are viable (Pasini 2004; Donohoe 1999).

If conditions necessary for proper placement of H3K27me3 are in fact required to antagonize placement of DNA methylation on the maternal DMD of *Rasgrf1*, then DNA methylation at the DMD will increase in the absence of SUZ12 and YY1. Because DNA methylation at the *Rasgrf1* DMD is normally restricted to the paternal allele, which is completely methylated, any increase in DNA methylation would arise on the maternal allele. To monitor the level of

Rasgrf1 DMD methylation in SUZ12- and YY1-deficient materials, we used COBRA (Xiong 1997). This involved treating DNAs with bisulfite and amplifying them using primers not overlapping with CpG dinucleotides, which will amplify templates without bias for either methylation state. We then digested the PCR products with BstUI. Methylated templates will retain the BstUI recognition site (CGCG) after amplification and will be digested, whereas unmethylated templates that underwent bisulfite conversion of either CG in the recognition site to TG will resist digestion. There should be an equal amount of digested and undigested PCR product when the maternal allele is completely unmethylated and the paternal allele is completely methylated. This is what we saw in embryoid bodies and blastocysts that were heterozygous respectively for the *Suz12* and *Yy1* mutations. This indicated that our COBRA assays accurately reported the presence of both methylated and unmethylated templates expected in these *Suz12* and *Yy1* expressing materials; however, it is not clear why *Suz12* heterozygous ES cells did not show this pattern. When we performed COBRA analysis on SUZ12-deficient embryoid bodies (EB) that had differentiated for six (P6) or nine (P9) days in vitro (Figure III.1.6A, III.1.B) or on trophoblast outgrowths from YY1-deficient blastocysts (Figure III.1.6C), we found a dramatic increase in the levels of the digested PCR product in three out of four samples of *Suz12 –/–* material and in the *Yy1 –/–* material, indicating that loss of SUZ12 or YY1 resulted in increased *Rasgrf1* DMD methylation. The near complete acquisition of DNA methylation in P9 EB lacking SUZ12 was confirmed by bisulfite sequencing (Figure III.1.6B lower panel), whereas unmethylated DNA was present in EB with a single functional allele of *Suz12*, though it is possible there is a quantitative increase in *Rasgrf1* DNA methylation when only one functional

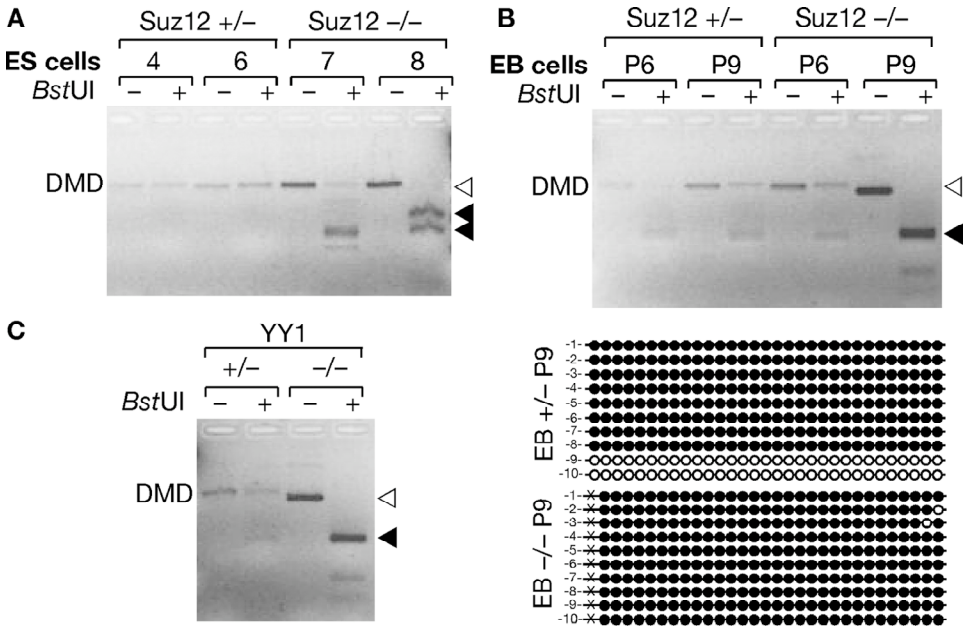


Figure III.1.6. Loss of H3K27 methylation potential enables inappropriate DNA methylation. (A, B) DNA methylation analysis at *Rasgrf1* in *Suz12 +/–* or –/– ES cells (A) and embryoid bodies (EB, B). EB were prepared from ES cells after 6 (P6) or 9 (P9) days in culture. Methylation analysis was by COBRA (A) and (B, top) and by bisulfite sequencing of P9 EB DNA (B, bottom). (C) COBRA analysis of trophoblast outgrowths from embryos lacking YY1, a CTCF and PRC2 co-factor. In COBRA, bisulfite treated DNAs were amplified using the core DMD spanning primer pair from D5 in Figure 1, and PCR products were left undigested (–) or digested with BstUI (+) prior to electrophoresis. Six BstUI sites are in the amplicon. DNA methylation preserves the sites in bisulfite treated DNA, whereas the sites are lost in unmethylated DNA. When one allele is fully methylated and the other unmethylated, COBRA will produce an equal quantity of uncut and cut bands after BstUI digestion. When the maternal allele acquires DNA methylation, amounts of digested products will increase; when the paternal allele loses DNA methylation, fewer digested products appear. Open triangles, unmethylated DNA; filled triangles, methylated DNA.

copy of *Suz12* is present. We do not know why only three out of four of the *Suz12 –/–* DNAs show hypermethylation. This could be an artifact of cell cultures, which can exhibit frequent and cyclic changes in DNA methylation (Kangaspeska 2008). Also, mutation of Eed, another component of PRC2, is known to cause hypermethylation and hypomethylation simultaneously, depending upon which CpGs are queried (Mager 2003). Given these precedents, it is possible that the eight CpGs we assayed in the BstUI sites are predominantly hypermethylated in cultured cells lacking SUZ12. Nonetheless, our data provide evidence that the antagonism between DNA and H3K27 methylation is reciprocal and that H3K27me3 antagonizes placement of DNA methylation. Furthermore, this mutual antagonism exists in at least three DNA sources: MEFs, embryoid bodies and trophoblast outgrowths.

We also explored the relationship between H3K9 and DNA methylation at *Rasgrf1*. H3K9 methylation has been strongly correlated with DNA methylation (Kouzarides 2007): Loss of the SUV39H1 and SUV39H2 H3K9 methyltransferases in mice simultaneously impairs accumulation of H3K9me3 across the genome (Peters 2001) and accumulation of DNA methylation at pericentric major satellite repeats, but not at minor satellite or C-type retroviral repeats (Lehnertz 2003). DNA methylation deficiencies were also noted in plants lacking H3K9 methyltransferases (Johnson 2002; Jackson 2002) with one study reporting that maintenance of DNA methylation was affected (Malagnac 2002). To investigate the relationship between H3K9me3 and DNA methylation at *Rasgrf1*, we asked if H3K9me3 controlled by SUV39H1 and SUV39H2 affected imprinted DNA methylation at *Rasgrf1*. To address this, we

performed methylation analysis on adult testes DNA using COBRA, bisulfite sequencing and a PCR assay that detected methylation status at a series of five HhaI sites in the DMD. Testes primarily contain cells of the germline, which will carry paternal epigenotypes, but some somatic cells are also present, which will carry both paternal and maternal epigenotypes. The COBRA analysis suggested that the DNA was hypomethylated in the SUV39H1- and SUV39H2-doubly deficient testes (Figure III.1.7A). When we measured the extent of DNA methylation using HhaI site-spanning Q-PCR assays, it was clear that the loss of *Rasgrf1* DNA methylation was significant (Figure III.1.7B). Bisulfite sequencing provided additional confirmation with higher resolution – there was a significant decrease in the number of DNA templates that were more than 80–100% methylated and an increase in the number that were 40–80% methylated in SUV39H1- and SUV39H2-doubly deficient testes (Figure III.1.7C) but there was no change in the abundance of DNAs that were completely unmethylated. The reduction in DNAs with the 80–100% methylated paternal epigenotype, and the increase in DNAs with the 40–80% methylated epigenotype suggests that SUV39H1 and SUV39H2 control the efficiency with which imprinted DNA methylation is established in mice. In Arabidopsis, the SUVH4 H3K9 methyltransferase is known to control maintenance of DNA methylation (Malagnac 2002).

III.1.5 Discussion

We report here the epigenetic states that exist within a 12 kbp interval centered on the *Rasgrf1* ICR. Both parental alleles were marked by DNA methylation in somatic tissue on a 1.4 kbp segment at the very 5′ end of this

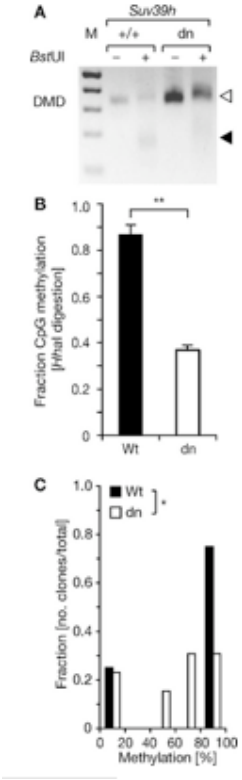


Figure III.1.7 Loss of Suv39h1 and Suv39h2 causes reductions in Rasgrf1 DNA methylation. (A) COBRA analysis of testes DNAs isolated from wild type (WT, +/+) mice or animals deficient for both Suv39h1 and Suv39h2 (dn), performed as described in Figure 6. (B) Methylation analysis using the PCR assay described in Figure 5B, middle panel, done in triplicate and analyzed by Q-PCR. A significant decrease of DNA methylation is observed in dn material (Student's t-test, $p<0.01$). (C) Bisulfite treated DNAs were amplified, cloned and sequenced. Data reported as the quintile distribution of methylation densities observed for the collection of clones (13 dn clones and 8 WT). The dn samples exhibit a significant excess of hypomethylated CpGs as assessed by likelihood ratio Chi-square analysis ($p<0.05$).

12,020 nt interval (D1–D2, Figure III.1.2C, III.1.2D). Downstream of this were segments that spanned the ICR that were paternally methylated in somatic DNA (D3–D8), and sperm DNA as well (D3–D5, Figure III.1.2B,III.1.2D). Not every CpG was assayed in this 12,020 interval, including those within the tandem repeats that constitute the DNA methylation programmer. H3K9me3 was present on both parental alleles at the core DMD immediately 5′ of the tandem repeats and within the ICR. H3K27me3 was present at this same location, but exclusively on the maternal allele. There was no appreciable dimethylation of these H3 residues at the core DMD (Figure III.1.4A, III.1.4B, III.1.4C).

The tandem repeats, consisting of approximately 40 copies of a 41 nt unit, influenced the placement of histone and DNA methylation (Figures, III.1.2B, III.1.2D, III.1.3 and III.1.4D) and can be considered a cis-acting methylation programming sequence, one of only a few naturally occurring ones known in mammals. Paternal allele DNA methylation was particularly sensitive to these tandem repeats, which control establishment of DNA methylation in the male germline at a 400 nt core DMD lying just 5′ of the repeats (Figure III.1.2B; Yoon 2002). The repeats also control spreading and maintenance of paternal allele DNA methylation in somatic tissue over a broader domain (Figure III.1.2B and III.1.2D; Holmes 2006).

Marking the core DMD with DNA methylation on the paternal allele and H3K27me3 on the maternal allele are coordinated and mutually exclusive events in wildtype cells with DNA methylation largely confined to the core DMD on the paternal allele and H3K27me3 on the maternal allele (Figures III.1.2D,

III.1.3B and III.1.4). The mutual exclusion arises because one epigenetic mark antagonizes the placement of the other. Five independent lines of evidence led to this conclusion. First, MEFs taken from mice lacking DNA methylation on the paternal DMD inappropriately accumulated H3K27me3 on the paternal allele (Figure III.1.4D). Second, MEFs treated with the DNA methyltransferase inhibitor, 5-azacytidine, accumulate elevated levels of H3K27me3 marks (Figure III.1.5A). Third, MEFs taken from mice with a maternally transmitted *Rasgrf1* ICR transgene that lacked DNA methylation had H3K27me3 on the transgenic DMD, whereas H3K27me3 was excluded by manipulations that inappropriately placed DNA methylation on the transgene (Figure III.1.5B, lower panel). Fourth, mutation of the *Suz12* component of PRC2, which is needed for activity of the EZH2 H3K27 methyltransferase in PRC2, ablated normal placement of H3K27me3 and enabled the maternal allele to inappropriately acquire DNA methylation (Figure III.1.6A, III.1.6B). Fifth, mutation of the *Yy1* gene, which is needed to recruit PRC2 to DNA and, like *Suz12*, is needed for effective placement of H3K27me3 also enabled the maternal allele to inappropriately acquire DNA methylation (Figure III.1.6C). Other studies have documented the cross-dependency of some histone modifications and DNA methylation (Tamaru 2003; Lehnertz 2003; Espada 2004; Jackson 2002; Tamaru 2001; Soppe 2002; Xin 2003; Tariq 2003; Ebbs 2006), and it has also been observed that H3K27me3 and DNA methylation can be mutually exclusive (Mathieu 2005). The studies described here provide evidence that H3K27me3 and DNA methylation are in fact mutually antagonizing epigenetic marks and that H3K27me3 facilitates allele-specific DNA methylation that exists at imprinted loci.

H3K9me3 was detected on both parental alleles indicating this mark is controlled differently from H3K27me3. However, it too participates in imprinted DNA methylation because the H3K9 methyltransferases, SUV39H1 and SUV39H2, are needed for optimal establishment of DNA methylation at the DMD in the male germline (Figure III.1.7).

We do not know how DNA and H3K27 methylation antagonize each other's placement; however, the literature highlights several molecular and developmental events, as well as protein factors that may be involved. Among these is the transcriptional state that is known to influence which of two mutually exclusive histone modifications is placed by the competing activities of polycomb (PcG) and trithorax (Trx) group proteins (Papp 2006). Additionally, differentiation state is known to influence genome wide epigenetic patterns in ES, MEF and neuronal progenitor cells (Bernstein 2006). At *Rasgrf1*, developmental stage also influences epigenetic states (Lees-Murdock 2003); the methylation programmer controls establishment of DNA methylation in the germline and maintenance in peri-implantation embryos (Yoon 2002; Holmes 2006) but not later in development. Interestingly, this same period is a critical interval for control of H3K27 methylation (Erhardt 2003). Finally, there may be a role for CTCF in the mutual exclusion of H3K27 and DNA methylation at *Rasgrf1*. CTCF and its binding sites have been shown to influence H19 DNA methylation (Pant 2003; Schoenherr 2003; Srivastava 2003; Fedoriw 2004; Pant 2004; Rand 2004) and CTCF binds at *Rasgrf1* as well (Yoon 2005). Genome-wide ChIP analysis identified locations enriched for CTCF (Kim 2007) and H3K27me3 (Mikkelsen 2007) in MEFs and Chi squared analysis reveals a significant co-localization of these marks at

imprinted versus non-imprinted loci (Table III.1.S2). This raises the possibility that, in addition to its role in preventing DNA methylation at other imprinted loci, CTCF helps to place H3K27me3 at *Rasgrf1*. CTCF functions in coordination with its binding partner, YY1, in activating the X chromosome (Donohoe 2007) and YY1 also inhibits DNA methylation at *Rasgrf1* (Figure III.1.6C), most likely through its ability to recruit PRC2 (Satijn 2001; Srinivasan 2005; Wilkinson 2006). Depending upon the consensus sequence considered, between one and twelve YY1 sites are predicted to lie within the DMD and repeat region (data not shown). Like CTCF, YY1 sites are enriched at other imprinted loci as well (Kim 2006). CTCF has additional binding partners including CHD8, which is associated with DNA methylation (Ishihara 2006). Using ChIP analysis, we could not detect CHD8 on either *Rasgrf1* allele (data not shown), suggesting that at *Rasgrf1*, other CTCF binding partners and functions might be more important, such as YY1.

Normal placement of DNA methylation on the paternal allele and H3K27me3 on the maternal allele both require the same tandemly repeated DNA sequence element, which we previously showed has DNA methylation programming activity (Yoon 2002; Holmes 2006; YJP, HH and PDS unpublished). However, DNA methylation is more rigidly dependent on the repeated sequence than are the histone modifications. Whereas DNA methylation on the paternal core DMD was typically completely lost when the repeats were deleted, H3K27me3 and H3K9me3 on the maternal DMD were respectively reduced to levels only 1/2 and 1/6 of those seen when the repeats were present. Repeated sequences have been shown to have methylation programming activity in other systems (Reinhart 2002; Chan 2006). Notably,

Table III.1.S2. Enhanced colocalization of CTCF and H3K27me3 at imprinted loci. Whole genome H3K27me3 ChIP data for imprinted and known genes in MEF cells were downloaded from http://www.broad.mit.edu/seq_platform/chip/ and experimentally verified CTCF site data were downloaded from http://insulatordb.utmem.edu/browse.php. After filtering the H3K27me3 ChIP data for sites with a read score of two or higher, the data sets were added as custom tracks on the UCSC Genome Browser and intersected in the intervals spanning 17,553 known genes and 53 imprinted gene regions. The intervals examined included the 100 kb 5′ of each gene (+100), sequences between the 5′ and 3′ ends of the genes (G), 100 kb 3′ of the genes (−100), and the entire stretch from 100 kb 5′ to 100 kb 3′ of each gene region (+100 to 100). The number of times H3K27me3 colocalized with CTCF in the indicated intervals is reported. The frequency of colocalization per kb was calculated for each interval examined, and the values for each of the known gene intervals were used to calculate an expected value for the corresponding imprinted gene intervals, given the total number of kbp in each of the imprinted gene intervals examined. The number of observed and expected colocalizations in the imprinted intervals was then used in Chi-square analysis.

Genes	Number of times H3K27me3 and CTCF colocalize in MEFs	Interval	Number of Kbp in interval	CTCF and H3K27me3 colocalized sites per kbp	Expected CTCF and H3K27me3 colocalized sites per imprinted gene	Chi square	P value
Known	60	+100	654506	9.17E-05	-	-	-
17,553	23	G	848864	2.74E-05			
	118	-100	1068023	1.10E-04	-	-	-
	201	+100 to -100	2563390	7.84E-05	-	-	-
Imprinted	8	+100	6997	-	6.41E-01	8.44E+01	4.01E-20
53	1	G	2981	-	8.15E-02	1.03E+01	1.30E-03
	12	-100	4195	-	4.64E-01	2.87E+02	2.09E-64
	21	+100 to -100	14173	-	1.11E+00	3.56E+02	2.18E-79

at the DM1 locus in humans, a repetitive element is associated with heterochromatin accumulation (Cho 2005). Interestingly, like the maternal *Rasgrf1* DMD and repeat sequences (Yoon 2005), the DM1 repeat also is a CTCF-binding insulator. CTCF appears to restrict the boundary of heterochromatinization at DM1, but it is not known if CTCF has a similar effect at *Rasgrf1*. Sequences with appreciable similarity to the *Rasgrf1* tandem repeats are not abundant in the mouse genome. However, the *Rasgrf1* repeat unit has striking similarity to the B repeat sequences on Xist (Figure III.1.S4). Because Xist RNA regulates placement of H3K27me3 on the inactive X chromosome and at an autosomal transgenic site in cis (Wutz 2002; Kohlmaier 2004), it is possible there is mechanistic overlap between epigenetic regulation by Xist and the *Rasgrf1* repeats.

We do not know what functional motifs enable the methylation programmer at *Rasgrf1* to control either DNA or H3 methylation. Its repeated nature may be sufficient (Chan 2006), possibly involving an RNA-dependent mechanism (Martienssen 2003). Other potentially important features include the CpG present in 36 of the 40 repeat units; GGGG tetramers that may facilitate the formation of G-quadruplex structures (Buge 2006), which in turn may alter the sensitivity of DNA to methyltransferase action (Hardin 1993); or CTCF sites known to lie in the *Rasgrf1* methylation programmer (Yoon 2005). BORIS, the male germline paralog of CTCF (Loukinov 2002), may also be important for function of the *Rasgrf1* methylation programmer.

Figure III.1.8 describes a model for the placement of DNA methylation and H3K27me3 in response to the *Rasgrf1* methylation programmer, their

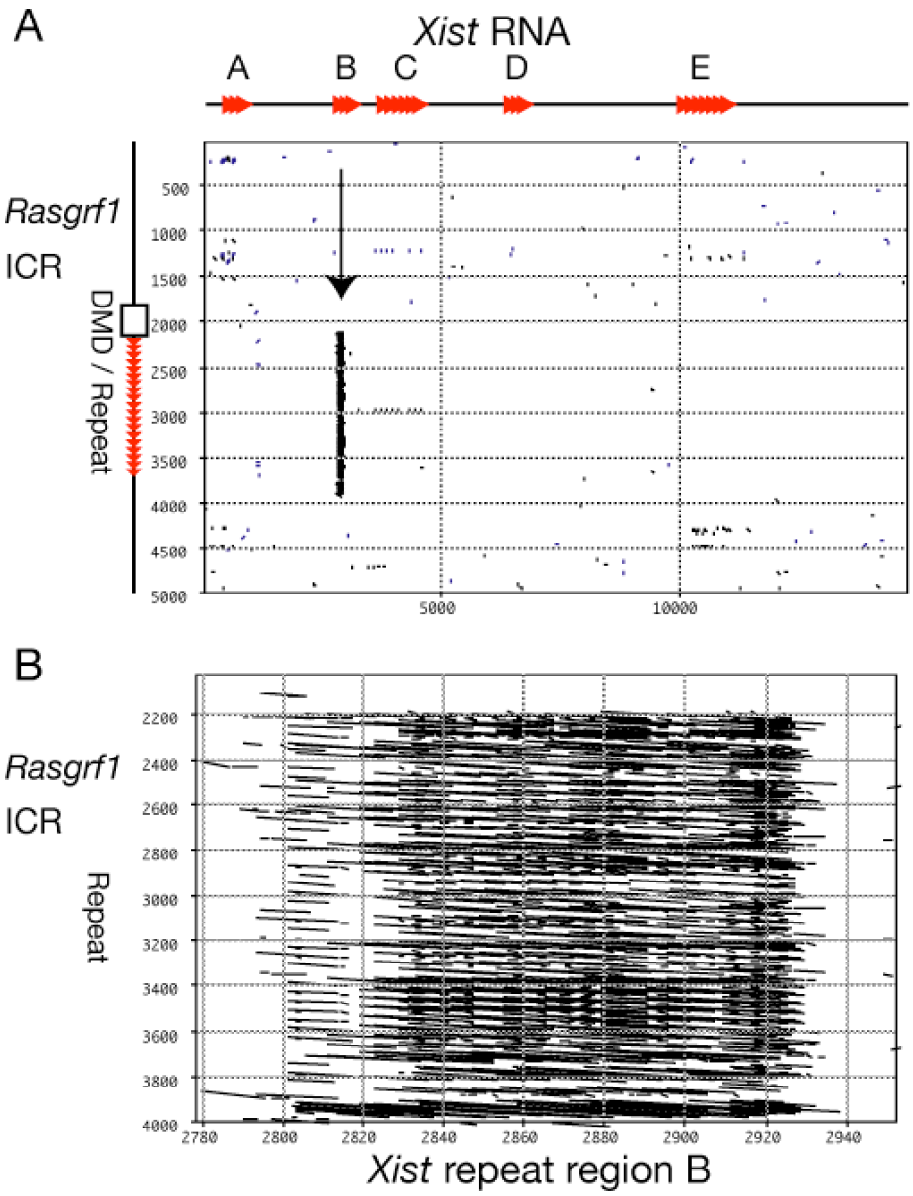


Figure III.1.S4. Dot plot of Xist and the *Rasgrf1* ICR. (A) Xist sequences, including the A, B, C, D and E repeats (17 kb) and *Rasgrf1* sequences including the DMD and repeats (5 kb) were aligned in a dot plot matrix. (B) Detail of the dot plot matrix in A that includes the Xist B element and the *Rasgrf1* repeats.

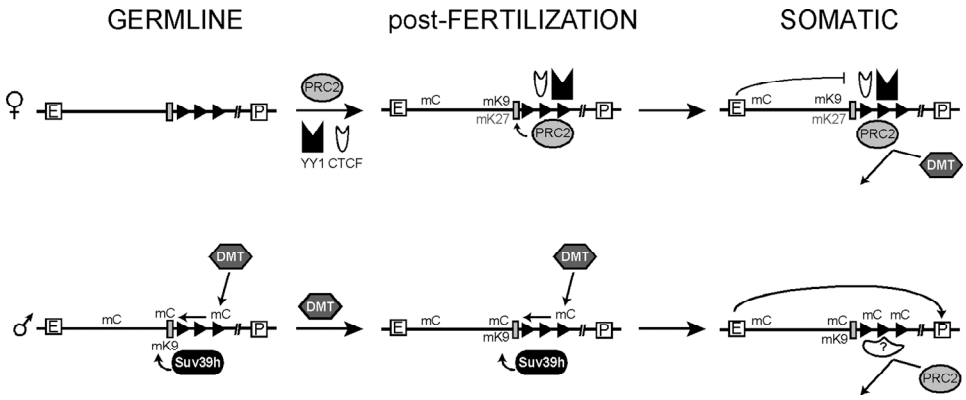


Figure III.1.8. Model summarizing epigenetic control at the ICR of *Rasgrf1*. The maternal allele (top) recruits YY1, PRC2 components, and possibly CTCF early during development or gametogenesis. H3K27 and H3K9 are both methylated (mK27 and mK9 respectively) in the vicinity of the core DMD (grey box), with optimal methylation depending upon the tandem repeats (rightward pointing black triangles, which constitute the methylation programmer). Once placed, H3K27me3 can exclude DNA methylation. On the paternal allele (bottom), the methylation programmer is active in the germline where it directs DNA methylation to the DMD (Yoon 2002) by a process optimized by SUV39H1 and SUV39H2 directed H3K9 methylation, and in the pre-implantation embryo where it maintains it (Holmes 2006). Once established in the germline, DNA methylation on the core DMD can expand to surrounding sequences, and exclude subsequent H3K27me3 during somatic development. In the neonatal brain, where *Rasgrf1* shows imprinted expression, recruitment of CTCF to the maternal allele allows the enhancer blocking activity of the DMD to silence the maternal allele by restricting interactions between a putative upstream enhancer (E) and the downstream promoter (P), while exclusion of CTCF by methylated DNA at the paternal allele allows expression.

antagonism, and the developmental timing of these events. However, it is unlikely that a universal rule dictates the regulation of DNA and H3K27 methylation at all loci within a species or among species. In human cell lines, some loci have been found at which DNA and H3K27 methylation occur simultaneously with one mark requiring the placement of the other (Vire 2005), whereas in Arabidopsis, DNA methylation does not seem to be closely associated with H3K27me3 (Zhang 2007) and in fact can be mutually exclusive (Mathieu 2005). Nonetheless, identifying the various rules that influence epigenetic programming of normal developmental states will provide insights for manipulating them for therapeutic benefit.

III.1.6 Materials and methods

III.1.6.1 Mouse strains and mutants

Mice used for DNA methylation analysis across the 12 kbp interval were F1 progeny of PWK and 129S4SvJae parents. Polymorphisms in these strains facilitated the assignment of a given clone from bisulfite PCR to one of the two parental alleles. Mice used to prepare MEFs for ChIP analysis across the 12 kbp interval were from strain 129S4SvJae backcrossed to C57BL/6 and included wildtype animals, animals carrying a repeat deletion (Yoon 2002) and animals containing an engineered polymorphism at the DMD that did not disrupt imprinting (Holmes 2006). All allele specific ChIP analyses were done using MEFs from mice carrying one of these mutations. Mice carrying the *Rasgrf1* ICR transgene will be described in a separate report (YJP, HH, PDS, in preparation). Previous reports describe the *Suz12* mutation and preparation

of homozygous ES cells and embryoid bodies (Pasini 2004) and the *Yy1* mutation and preparation of trophoblast outgrowths (Donohoe 1999).

III.1.6.2 Chromatin immunoprecipitation (ChIP) analysis

MEFs from 13.5 day old F1 embryos of C57BL/6 and 129S4Jae parents were used for ChIP analysis. MEF cells no older than passage four were grown in to confluence in DMEM with 10% fetal calf serum on 15 cm dishes. We cross linked chromatin with formaldehyde by adding 675ml of a 37% solution (1% final concentration) directly to 25 ml of culture medium and rotating cells lightly at room temperature for 10 min. The cross linking was quenched by adding 1 ml 2.5 M glycine (final 100 mM) and incubating for an additional 10 min. Cells were then rinsed with phosphate buffered saline containing proteinase inhibitor mix (PBS complete), used according to the manufacturer's instructions (Roche, #11697498001). Cells were then scraped into 2 ml PBS complete, centrifuged at 4000xg for 5 minutes, and washed again with PBS complete. Cells were resuspended in lysis buffer (1% SDS, 10mM EDTA and 50mM Tris, pH 8.1 supplemented with proteinase inhibitors) to a concentration of 1-1.5x10^7 cells /ml and frozen in 1ml-aliquots at –80°C for future use. For immunoprecipitations, we thawed frozen cells, added fresh proteinase inhibitors, supplemented with Pepstatin A (0.7mg/ml) and sonicated samples on ice until the DNA was sheared to ~500 bp. Chromatin was then diluted to 10 mls in dilution buffer (0.01% SDS, 1.1% Triton X- 100, 1.2mM EDTA, 16.7mM Tris-HCl, pH 8.1, 167mM NaCl supplemented with proteinase inhibitors). Two ml aliquots were used for each precipitation and for no antibody and input controls. Lysates were precleared with 60ml protein-G-agarose beads coated with SS-DNA and BSA (Millipore #16-201), for 1hr at

4°C followed by centrifugation at 2500xg, 1 min. Supernatants were transferred to new tube and desired antibody added (5-10ml from a 100mg/ml stock) and samples were rotated over night at 4°C. The next day, protein-G-agarose beads were added to samples (except input controls) which were rotated for 2h at 4 °C followed by centrifugation at 2500xg, 1 min. After discarding supernatant, we washed the beads twice in low salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 150mM NaCl), twice in high salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 500mM NaCl) once in LiCl buffer (0.25M LiCl, 1% IGEPAL CA630, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris, pH 8.1) and twice in TE (10mM Tris pH 8.8, 1mM EDTA). Each wash was for 5 minutes with rotation at 4°C. Chromatin was eluted from beads using 200ml elution buffer (1% SDS, 0.1M $NaHCO_3$) twice at 65°C and we pooled the eluted materials. We then reversed the cross links by adding 16ml 5M NaCl and incubating samples at 65°C for 4-6 h. We did this as well with material reserved for the input control. Protein was removed by adding 20ml 1M Tris (pH6.5), 10μl 0.5 M EDTA, and 1-2ml 10mg/ml Proteinase K followed by a 1h incubation at 45°C and an extraction with phenol/chloroform. DNA was transferred to new tubes containing 2-5ml Glycoblue (Ambion, #9516) and precipitated by adding 2 volumes ethanol. We ran Q-PCR on DNA after dissolving it in 50ml 10 mM Tris, pH 8.5.

Modified histone-specific antibodies were from Millipore/Upstate (H3K9me2 item 07-441 lot 29698, H3K9me3 item 07-442 lot 24416, H3K27me2 item 07-452 lot 24461, H3K27me3 item 07-449 lot 24440) and Thomas Jenuwein, IMP, Austria (H3K9me3) (Peters 2003). Specificity of

antibody from Thomas Jenuwein's lab has been reported (Peters 2005). Validations of commercial antibody specificities are publicly available from the manufacturer (see http://www.millipore.com for certificates of analysis for each catalog item and lot number). These used a combination of methods including immunoblots of membranes containing an array of synthetic methylated histone peptides and solution binding assays. Additionally, immunoblots of membranes containing mammalian histones were used in competition tests that assayed the ability of immunogens and related peptides with distinct sites or numbers of methyl modifications to compete for antibody binding. Anti H3K27me3 binding to mammalian histones was modestly competed by a peptide containing dimethylated K27, however, no H3K27me2 was detected at *Rasgrf1* (Figure III.1.4) indicating any recognition of H3K27me2 by antibody designed to recognize H3K27me3 did not confound our assays at *Rasgrf1*. Note that in our hands the commercial antibodies produced similar results as those obtained by others using independently prepared noncommercial antibodies with the same specificities (Figure III.1.S2).

Validations of commercial antibody specificities are publicly available from the manufacturer (see http://www.millipore.com for certificates of analysis for each catalog item and lot number). The DNA recovered after ChIP was used for Q-PCR with input chromatin and mock immunoprecipitations without antibody serving as controls. Q-PCR was performed in triplicate with SYBR green detection using primers listed in Table III.1.S3. Ratios of bound to input signals are reported.

III.1.6.3 DNA methylation analysis

On the day of bisulfite treatment, we prepared fresh 10M NaOH, 6.3M NaOH, 20mM hydroquinone (Sigma H9003) and 3.9M bisulfite solution. The bisulfite solution included 16.2g of Na bisulfite in 30ml dH2O, pH adjusted to 5.1 with 10M NaOH. Then we added 1.32ml of 20mM hydroquinone and adjusted the total volume into 40ml, sterilized the solution with a 0.2mm filter and stored protected from light. We used 2mg of genomic DNA, sheared by passage through a 30g needle 15 times, and purified by phenol/chloroform extraction, followed by alcohol precipitation. We resuspended DNA in 20ml dH2O and transferred to a 0.65ml tube, heated at 95°C for 5min and quenched on ice. Then we added 1ml of 6.3M NaOH, incubated at 39°C for 30min, followed by 208ml bisulfite solution, mixed and incubated the reaction at 55°C for 16hr in the dark, heating to 95°C for 5min every 3 hours. At the end of the incubation, we desalted samples using a Qiagen Quickspin column, following the manufacturers instructions, and eluted the sample in 50ml EB (Qiagen). We completed the reaction by adding 2.5ml of 6.3M NaOH (final 0.3M) and incubated at 37°C for 15 min. Finally, we purified DNA samples using Qiagen Quickspin columns, after adjusting the pH to 5 using 3M Na acetate, and eluted DNA with 30ml of warm EB, allowing the EB to stand in the column for 1-2 min. We used 2ml DNA for PCR and cloned the products into TOPO pCR2.1 (Invitrogen) using the manufacturers instructions. Primers used for PCR of bisulfite treated DNAs are listed in Table III.1.S3. PCR was done with the addition of 1.5 M betaine and 5% DMSO to enhance the yield in PCR of AT-rich, converted DNA. ExTaq HS DNA polymerase (Takara, Japan) was used for hotstart PCR. The bisulfite converted and amplified DNA was either cloned and sequenced or subjected to COBRA (Xiong 1997) using

Table III.1.S3. Primers used for PCR amplification.

DNA Amplified	Primer Pair	Lab code
Rasgrf1^{tm1Pds} Repeat deletion	5'-CACTTCGCTACCGTTTCGC-3' 5'-TGTCCTCCACCCCTCCACC-3'	P3 PDS16 P4 PDS17
C1	5'-GCCCTCCCTGTCTGTATGAG-3' 5'-TGTCAAACCCCCAAACTAGC-3'	PDS104 PDS105
C2	5'-ATACGTGCACACAGGCAAAA-3' 5'-GGCCTGGCATAGTAGCAGAA-3'	PDS650 PDS651
C3 (Core DMD)	5'-CACATCCATCCGTGGCTACCGCTATTGCTGT-3' 5'-GCGAAGTGCGGCAGCAGCAGCGA-3'	P1 PDS12 P2 PDS13
C4	5'-GGGTGGGTCTTTCTTTGTCA-3' 5'-CAACCGATAAGCCAAGTAGGA-3'	PDS648 PDS649
C5	5'-AACAATTCTGGCTCCCACAG-3' 5'-CCTCCCTCCAAAAGGACACT-3'	PDS108 PDS109
C6	5'-TAGTGATCCCCTTGCCTTTG-3' 5'-CCATATCACACCCTGGCTCT-3'	PDS106 PDS107
D1	5'- TGTTGTGTTATGGTTTATATATGGAGGTTAGAG-3' 5'-ACACCTAAAACCCATACAACTATTTCCCTAATA-3'	PDS629-Y PDS630
D2	5'-TTAGTGATTGTGATGATTTTTTGTTTGAGTT-3' 5'-TTAAAATTCTATCAAACCCCCAAACTAACTAC-3'	PDS368 PDS369
D3	5'-AGTGTATTGTGTTTTTATTGGTTATTTTAAAGGATAGAAT-3' 5'-AAACCATCACAAAAAACCACACAACTC-3'	PDS559 PDS356
D4	5'-GGGATTTAAAATGTTTTTTTTTGGTTATTAGGGAT-3' 5'-ACATTCTCAACAAAAACAATAACCTACCTA-3'	PDS269 PDS270
D5	5'-GGAATTTTGGGGATTTTTTAGAGAGTTTATAAAGT-3' 5'-CAAAAACAACAATAATAACAAAAACAAAAACAATAT-3'	PDS271 PDS272
D5 (RepΔ)	5'-GGAATTTTGGGGATTTTTTAGAGAGTTTATAAAGT-3' 5'-CTATATTAAATCCTTTTATCCACTATCCTCCACCC-3'	PDS271 PDS287
D6	5'-TAGTTGGAGATATTTTGATGAGGAAGATTAGATTTG-3' 5'-AACCATCCTAATTAACAAAACAAAACCC-3'	PDS623 PDS4
D7	5'-AAGGTATGTGAATTTATATGTGGTTGGGAA-3' 5'-TCCATTCCTCCCTCCAAAAAAACAC-3'	PDS563 PDS564
D8	5'-TGGGAGGAAGGATTGTGTATATATGGAT-3' 5'-ACTTCCAAAACACTCTCTCTACTTTCTCTA-3'	PDS275 PDS276
Rpl32	5'-CATGCACACAAGCCATCTACTCA-3' 5'-TGCTCACAATGTGTCCTCTAAGAAC-3'	PDS72 PDS73
Actin	5'-CAGTTCGCCATGGATGACGATATCG-3' 5'-CCGCGAAGCCGGCTTTGCACATG-3'	PDS38 PDS39
Transgenic DMD	5'-CACATCCATCCGTGGCTACCGCTATTGCTGT-3' 5'-CCTGCAGGTCGACATAACTTC-3'	pYJC6F2 pYJC6R2
Charlie	5'—TTGAGAATCGGATGGGAGAC-3' 5'-AAGAACTGTCTTATTCAGGC-3'	PDS660 PDS661
Hoxa9	5'-ACCGACTCTGCCAGCTTTAC-3' 5'-TCTCCCTTCTCAAACCCTCA-3'	PDS646 PDS647
COBRA Primers	5'-AGAGAGTATGTAAAGTTAGAGTTGTGTTGTTG-3' 5'-CAAAAACAACAATAATAACAAAAACAAAAACAATAT-3	PDS225 PDS272

BstUI digestions. In this assay, cytosine methylation enables digestion, whereas absence of methylation prevents it. Note for COBRA analysis of P9 EB from materials carrying *Suz12* and *Yy1* mutations, we used a semi-nested PCR assay for region D5 (see Figure 1) with primers 225/272 in the first round and 271/272 in the second round. Assays for DNA methylation using HhaI digested DNAs were as described previously (Yoon 2002).

In order to determine the DNA methylation state using bisulfite sequencing shown in Figure 2, we used F1 progeny of PWK and 129S4Jae parents. We sequenced DNA from the inbred parents in the vicinity of regions D1-D8 and identified indels and SNPs in regions D1, D2, D4, D5, D7 and D8. When analyzing a given bisulfite sequence from the F1 DNA, we could readily determine the parent of origin by determining which indels or SNPs were present. Note that C⇔T polymorphisms were uninformative because bisulfite treatment creates T residues from C residues in the PCR product.

III.1.8 Acknowledgements

Barna Fodor, Susanne Opravil and Thomas Jenuwein (IMP, Vienna, Austria) for sharing antibodies and the Suv39h1/2 double null mutant material; Hisato Kobayashi and Hiroyuki Sasaki (National Institute of Genetics, Mishima, Japan) for sharing unpublished data and primer sequences; Mitsuyoshi Nakao (Dept. of Regenerative Medicine, Kumamoto University, Japan) for sharing the ChIP grade CHD8 antibody and members of the Soloway lab for comments on the manuscript

III.1.9 Author contributions

Conceived and designed the experiments: AML YJP PDS. Performed the experiments: AML YJP CMM GAD JMP. Analyzed the data: AML YJP CMM PDS. Contributed reagents/materials/analysis tools: HH DP XM MED JTL KH. Wrote the paper: AML YJP CMM PDS.

III.2 SUFFICIENCY OF THE *RASGRF1* DNA REPEATS TO IMPART IMPRINTING TO AN ECTOPIC LOCUS

III.2.1 Abstract

DNA methylation is one of several classes of epigenetic modifications and has an important role in early development and X-inactivation, as well as in genomic imprinting. Proper DNA methylation is critical, as aberrant methylation patterns can lead to several human diseases, including cancer. However, the regulation of methylation is not well understood. Imprinted genes provide an ideal system for studying the regulation of DNA methylation for two reasons: they exhibit a predictable pattern of methylation and a predictable pattern of expression. Also, for four loci, *cis*-acting signals controlling DNA methylation have been identified: *Igf2/H19, Igf2r, Snrpn,* and *Rasgrf1*. At one such locus, *Rasgrf1,* a 41 bp element repeated 40 times has been identified as necessary for methylation of a differentially methylated domain (DMD), which is needed for proper imprinted expression. The experiments proposed here test whether the *Rasgrf1* DMD and repeats are sufficient to impart imprinting to an endogenous non-imprinted locus using the mouse as a model system. Utilizing a targeted knock-in approach, the *Rasgrf1* DMD and repeats were inserted into the non-imprinted *Wnt1* locus in order to test the ability of the DMD and repeats to convert *Wnt1* to an imprinted locus using allele-specific methylation and expression assays. Unfortunately, blastocyst microinjection did not yield germline chimeric animals.

III.2.2 Introduction

Using imprinted gene systems, several groups have demonstrated that there are *cis*-acting signals that prevent the genome from becoming uniformly methylated. For example, at the *H19/Igf2* locus, perhaps the best studied of the ~100 known mouse imprinted genes, a targeted deletion experiment identified a DMD that is necessary for maintenance of allele-specific methylation. Paternal allele DNA methylation marks were correctly established, as evidenced by analysis of sperm DNA, but were not properly maintained throughout development (Thorvaldsen 1998). Further studies, using RNAi against CTCF or mutated CTCF sites within the DMD, showed that the presence of CTCF binding sites is necessary for maintenance of the unmethylated state on the maternal allele (Fedoriw 2004; Schoenherr 2003). However, the sequences that establish proper DNA methylation at this locus remain unknown. At two other well-studied imprinted genes, *Igf2r* and *Snrpn*, two important *cis*-acting regulatory signals have been identified. The 6-12 bp allele discriminating signal (ADS) prevents paternal allele methylation, while the 7-8 bp *de novo* signal (DNS) establishes methylation in the female germline (Wutz 1997; Birger 1999; Kantor 2004). However, these sequences have only been tested at ectopic sites in the genome. Therefore, their function at the endogenous *Igf2r* and *Snrpn* loci has not been confirmed**.**

In addition to those at *H19/Igf2* and at *Igf2r*, *cis*-acting regulatory signals have been identified at a second imprinted locus, *Rasgrf1* (RAS guanine nucleotide releasing factor). *RASGRF1* is a RAS activator on mouse chromosome 9 and a homologue of the yeast gene *cdc25* (Cen 1993). *Rasgrf1* was identified as an imprinted gene by restriction landmark genome

scanning (RLGS), a technique that identifies DNA sequences that are subject to allele-specific methylation (Plass 1996). In mice, the *Rasgrf1* paternal allele is exclusively methylated at a differentially methylated domain (DMD) in all tissues and is expressed exclusively from the paternal allele in neonatal brain.

The *Rasgrf1* locus is associated with a repeated DNA sequence element, a feature common to many imprinted genes (Shibata 1998). The DNA repeat element (repeats) is a 41 bp repeat unit present in 40 copies and is located directly 3' of the DMD (Pearsall 1999). Our lab has used targeted knock-out mice to demonstrate that the repeats are necessary for imprinting at the endogenous *Rasgrf1* locus. Paternal inheritance of the repeat deletion leads to defects in proper establishment of methylation in the male germline and to defects in proper maintenance of methylation in the pre-implantation embryo (Yoon 2002; Holmes 2006). Maternal inheritance of the repeat deletion does not affect the methylation or the expression status of either the maternal allele in *cis*, or the paternal allele in *trans* (Yoon 2002). However, paternal inheritance of the repeat deletion results in a loss of establishment of methylation at the DMD, which leads to silencing of the normally expressed paternal allele (Yoon 2002). Similarly, *Zp3Cre* mediated deletion of the repeats at e0.0, after establishment of DMD methylation, results in a failure to maintain DMD methylation (Holmes 2006). Conversly, DNA methylation is properly maintained after *Meox2Cre* mediated deletion at e5.5, defining a critical period for the requirement of the DNA repeats for proper DNA methylation at *Rasgrf1* (Holmes 2006).

Furthermore, enhancer-blocking assays indicate that the DMD is able to

act as a methylation-sensitive enhancer blocker when placed between an enhancer and a promoter driving the neomycin reporter (Yoon 2005). Gel shift and chromatin immunoprecipitation (ChIP) assays showed that CTCF binds the DMD but, when the CTCF binding sites are mutated, the DMD is unable to act as an enhancer blocker (Yoon 2005). Therefore, the DMD and repeats constitute a *cis*-acting binary switch that regulates DNA methylation and imprinted expression at *Rasgrf1*.

In the current model of imprinting at *Rasgrf1*, the repeats establish paternal allele-specific DNA methylation at CpG dinucleotides in the DMD, preventing CTCF binding and allowing promoter-enhancer interaction, resulting in transcriptional activation (Figure III.2.1; Yoon 2005). The maternal allele is unmethylated at CpG dinucleotides in the DMD, allowing binding of CTCF to the DMD and preventing the *Rasgrf1* enhancer from interacting with the promoter, resulting in suppression of transcription (Yoon 2005). *Rasgrf1* is one of a handful of loci for which the *cis*-acting methylation regulatory elements have been identified and is the **ONLY** paternally expressed locus for which these sequences are known. Additionally, more is known about *cis*-acting control of both imprinted methylation and imprinted expression of *Rasgrf1* than for any other imprinted locus. For this reason, our lab was uniquely positioned to attempt to reconstitute imprinting from the *cis*-acting *Rasgrf1* DMD and repeats.

The DMD and repeats elements were inserted into the *Wnt1* locus to determine whether these sequences are sufficient to impart imprinting to a non-imprinted locus. The *Wnt1* locus was selected for several reasons. First,

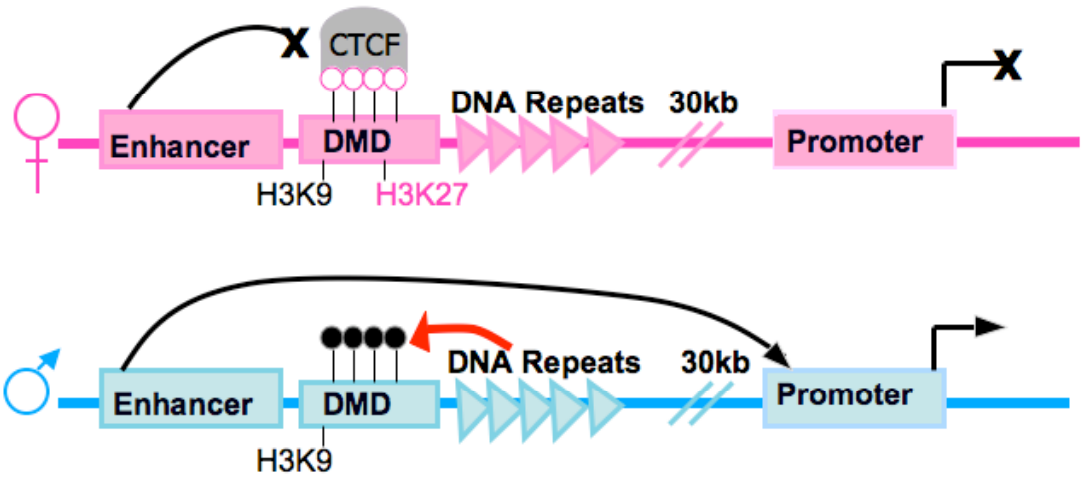


Figure III.2.1. Current model for epigenetic control of imprinted expression at *Rasgrf1.* One the paternal allele (depicted in blue), a series of DNA repeats are responsible for the establishment and maintenance of paternal allele-specific DNA methylation at CpG dinucleotides in the DMD (Yoon 2002; Holmes 2006). Placement of DNA methylation at the DMD prevents CTCF binding and allowing promoter-enhancer interaction, resulting in transcriptional activation (Yoon 2005). Placement of DNA methylation at the DMD also blocks the binding of H3K27me3 (Lindroth 2008). The maternal allele (depicted in pink) is unmethylated at CpG dinucleotides in the DMD. The unmethylated status if the DMD allows binding of CTCF and prevents the *Rasgrf1* upstream enhancer from interacting with the promoter, resulting in suppression of transcription (Yoon 2005). Lack of methylation at the DMD also allows placement of the silencing mark H3K27me3 (Lindroth 2008). H3K9me3 is found on both the maternal and the paternal allele and therefore does not appear to control differential expression. Circles above the DMD represent the methylation status of CpG dinucleotides within the DMD. Black circles represent a methylated DMD and white circles represent an unmethylated DMD.

it is expressed in neonatal brain, where the *Rasgrf1* DMD and repeats are known to control *Rasgrf1* expression, minimizing concerns about tissue specificity (Wilkinson 1987). Also, the *Wnt1* enhancer has been characterized thoroughly, eliminating the need to estimate enhancer boundaries (Rowitch 1998). In addition, animals with one expressed *Wnt1* allele are viable and fertile, eliminating lethality or sterility as concerns in propagating transgenic mice (Rowitch 1998; McMahon 1990). Most importantly, gene targeting experiments have shown that the *Wnt1* enhancer is absolutely required for *Wnt1* expression, that there are no additional enhancers in the region that could allow *Wnt1* expression, and that insertion of sequences up to 3 kb between the promoter and enhancer does not result in positional-dependent silencing (Danelian 1997; Echelard 1994).

If successful, this would have been the first example of converting a gene from a non-imprinted to an imprinted state at its endogenous position in the genome. Tanimoto *et al*. had success with insertion of the *H19* differentially methylated region (DMR) into the human β-globin locus; however, this was done using a transgene system, not at an endogenous locus in the genome (Tanimoto 2005). Similarly, our lab has created a transgene with the *Rasgrf1* DMD and repeats between the human A-γ globin promoter and the β-globin enhancer. Proper methylation patterns were observed upon both paternal and maternal inheritance of this transgene (Herman, unpublished). However, passage of the transgene first through the paternal germline and then through the maternal germline, resulted in failure of maternal reprogramming (Figure III.2.2) (Herman, unpublished). That is, the paternal methylation marks laid down when passed through the paternal germline were

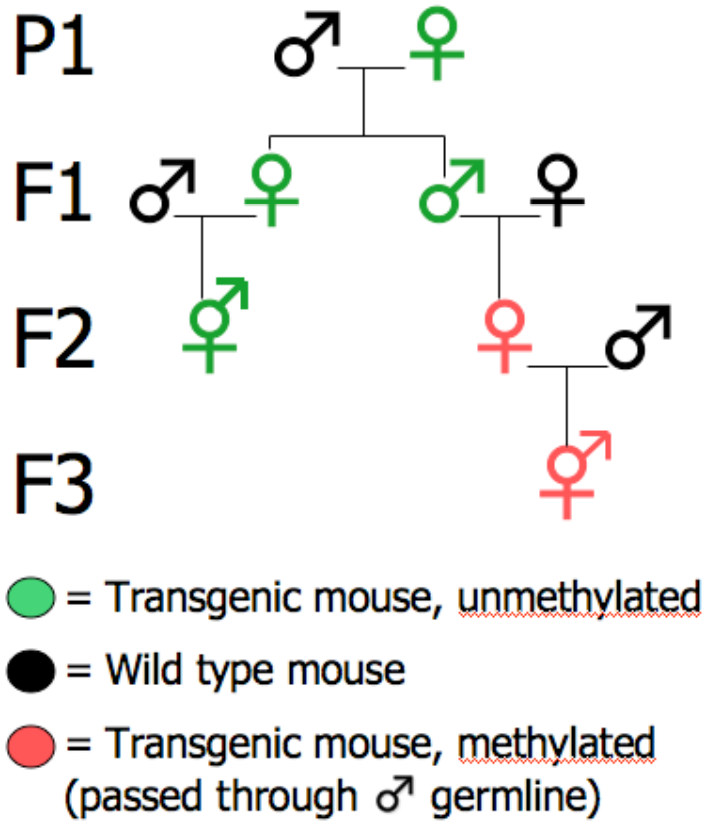


Figure III.2.2. Pedigree and methylation status of a minimal *Rasgrf1* transgene. A minimal *Rasgrf1* transgene containing the DMD and repeats between the human A-γ globin promoter and the β-globin enhancer was passed through both the male and female germlines. Proper DNA methylation was observed upon inheritance of the transgene from the male and female founders (Herman, unpublished). However, passage of the transgene first through the paternal germline (F1 ♂) and then through the maternal germline (F2 ♀), resulted in failure of maternal reprogramming (F3 progeny) (Herman, unpublished) As the paternal methylation marks were not removed upon passage through the maternal germline, this could indicate that, in addition to the DMD and repeats, a maternal demethylation signal is needed for proper imprinting at *Rasgrf1*. Male (♂) and female (♀) symbols are color coded to represent wildtype animals (black), methylated transgenic animals (green), and unmethylated transgenic animals (red). P1 is the founder generation, F1 is the first filial generation, F2 is the second filial generation, and F3 is the third filial generation.

not removed upon passage through the maternal germline. This indicated that either additional sequence features might be required for proper germline reprogramming or the aberrant DNA methylation was the result of a transgene artifact. Transgene artifacts are not uncommon among imprinted transgenes, as the majority fail to undergo proper maternal reprogramming after passage through the paternal germline (Bartolomei 1997). Work done by Dr. Yoon Jung Park determined that proper germline reprogramming did occur, but that somatic maintenance failed (discussed in Appendix Chapter III.6), leaving open the possibility that sequences important for somatic maintenance are missing from the transgene. Nevertheless, these experiments have demonstrated that it is possible to impart imprinting to a non-imprinted locus and that the experiments proposed here are not only informative, but also feasible.

The experiments described here intended to further current understanding of the *cis*-acting signals that regulate DNA methylation and proper imprinting, focusing on the signals at one gene, *Rasgrf1*. *Rasgrf1* is useful for studying the regulation of DNA methylation, since it is one of a handful of imprinted loci whose *cis*-acting signals controlling DNA methylation have been identified. Using a knock-in approach, I have attempted to address whether the *Rasgrf1* DMD and repeats are sufficient to impart imprinting to another locus

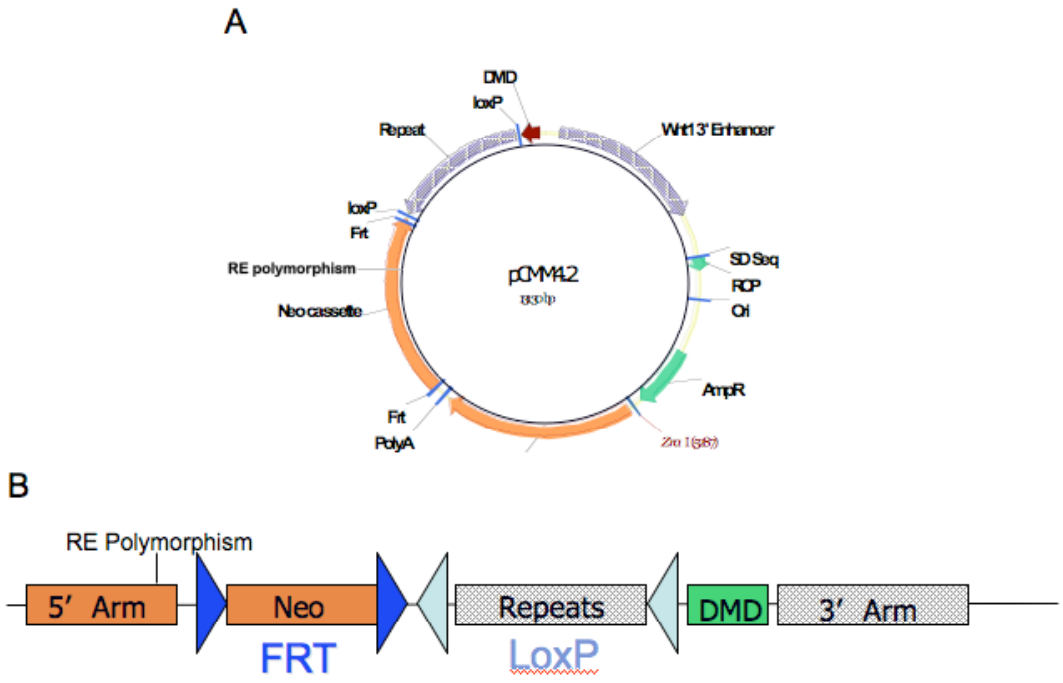


Figure III.2.3. *Wnt1* targeting vector design. (A) The plasmid containing the *Wnt1* knock-in vector consists of a 5' homologous arm containing 2.6 kb of the Wnt1 coding sequence, the *Rasgrf1* DMD, the *Rasgrf1* repeats flanked by LoxP sites, a neomycin resistance cassette flanked by FRT sites, and a 3' homologous arm containing 4.8 kb of the *Wnt1* 3' enhancer. The *neo* selectable marker is flanked by FRT sites, which will allow for later removal of the *neo* cassette. Likewise, the repeats are flanked by loxP sites and can be removed by utilizing the loxP/*Cre* system. (B) The plasmid was digested with ZraI to yield the linearized product used for electroporation.

III.2.3 Materials and methods

III.2.3.1 Targeting vector design

A knock-in vector was prepared which facilitated insertion of the *Rasgrf1* DMD and repeats between the *Wnt1* coding sequence and the *Wnt1* 3' enhancer. The targeting vector contains: a) a 5' homologous arm containing 2.6 kb of the Wnt1 coding sequence, b) the *Rasgrf1* DMD, c) the *Rasgrf1* repeats flanked by LoxP sites, d) a neomycin resistance cassette flanked by FRT sites, e) and a 3' homologous arm containing 2.1 kb of the *Wnt1* 3' enhancer (Figure III.2.3). A negative selectable marker is not included, as increased vector size reduced vector stability, with the DNA repeats selectively deleted. The *neo* selectable is included to select for homologous integration after electroporation into embryonic stem (ES) cells. Due to the presence of flanking FRT sites, *neo* can be removed after targeting to eliminate any effect of the associated promoter. Likewise, the repeats can be removed by utilizing the loxP/*Cre* system. The loxP-flanked repeats havebeen used previously by our lab and function as the wildtype repeats in that they allow paternal allele-specific methylation of the DMD and paternal allele-specific expression (Holmes 2006). By crossing knock-in mice to *Cre*-expressing mice, the repeats can be deleted at specific developmental stages. This feature of the targeting vector will confirm that the repeat elements are necessary for establishment and maintenance of DNA methylation at the knock-in DMD, and that the repeats and the DMD function together to produce imprinted expression.

III.2.3.2 ES cell targeting

The targeting vector was linearized with *ZraI* and electroporated into passage 8 v6.5 male (B6x129 F1) ES cells. Electroporation conditions were 240v, 250uF on a Bio-Rad Gene Pulser Xcell machine. Electroporated ES cells were plated onto 10x10cm plates without G418 to recover. After one day, cells were switched to media with 200ug/mL G418 for selection.

III.2.3.3 Identification of targeted ES cell clones

ES colonies surviving G418 selection were picked to 24 well plates and allowed to grow. Each well was split into two fractions. One was frozen for later expansion, and one was re-plated for DNA extraction. Targeted ES cell colonies were identified via Southern blot after DNA extraction and digestion with *NdeI*. Using a 5' probe external to the targeting vector (amplified with PDS497 5' – GAA GTG GGG CAC ATC ATT – 3' and PDS492 5' – CAT TTG CAC TCT CGC ACA – 3'), wildtype cells produce an 11.47 kb band when digested with *NdeI*, while targeted cells produce a 9.43 kb band. The results of the 5' probe were confirmed with a 3' external probe (amplified with PDS349 5' – AAT ATG CCT GAC GCA CCT TC – 3' and PDS350 5' – CAC TTC TCT CTG GGC CTC AC – 3') producing a 7.47kb band for targeted cells.

III.2.3.4 Blastocyst microinjection

Successfully targeted cells were injected into either C57/BL6, CD1, or C57/BL6xCD1 F1 blastocysts and transplanted into a pseudopregnant recipient mother.

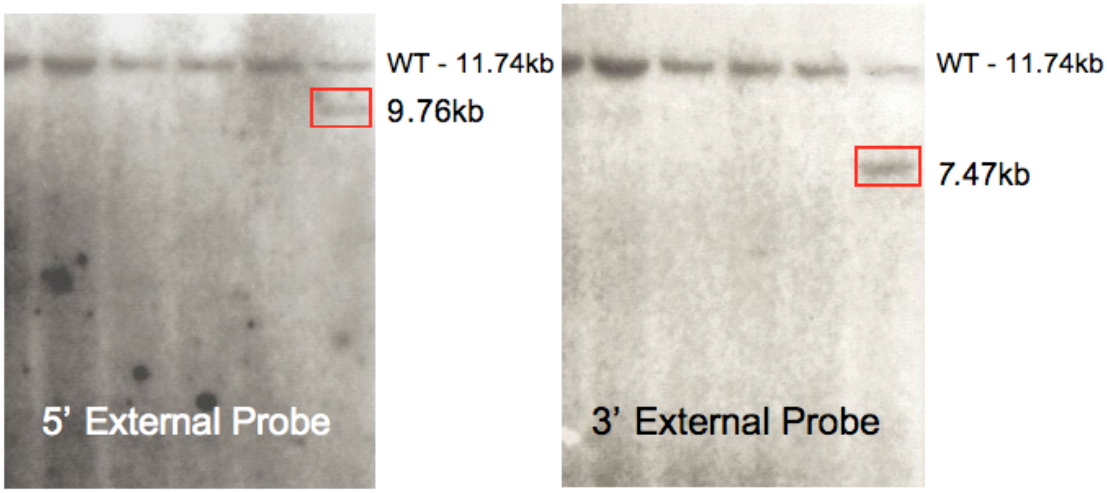


Figure III.2.4. Southern blot identification of targeted ES cells. Embryonic stem cell clones were digested with the restriction enzyme NdeI and were screened for evidence of homologous recombination with the targeting vector via Southern blot. Two different P^{32} labeled probes were used to confirm targeting. The 5' external probe yields an 11.47kb band corresponding to the wildtype allele. An additional smaller band of 9.76kb is seen if a correctly targeted allele is present. The 3' external probe also yields an 11.74kb band corresponding to the wildtype allele. A correctly targeted allele will result in a smaller 7.47kb band. In both the 5' and the 3' probe gels, the lane farthest to the right shows a banding pattern consistent with a targeting event.

III.2.3.5 Identification of chimeric progeny

Knock-in progeny were identified by PCR on tail DNA. PCR was done using primers PDS288 (5' – TTA CCC AGC TTC TCA TAG GCG C – 3') and PDS511 (5' – TCC CCT ACC CGG TAG AAT TGG A – 3') and was run under conditions 94C for 30 seconds, 59C for 30 seconds, 72C for 30 seconds for 40 cycles.

III.2.3.6 Identification of germline chimeric animals

Progeny carrying a knock-in allele were be bred to wildtype mice, and the resulting F1 generation was screened by PCR to identify founders that demonstrate germline inheritance.

III.2.4 Results

Electroporation of the pCMM4.2 targeting vector into v6.5 ES cells resulted in a single targeted ES cell colony in approximately 830 screened clones (Figure III.2.4), correlating to a targeting frequency of 0.12%. This single targeted ES cell clone was subsequently microinjected into either B6, CD1, or B6xCD1 F1 blastocysts. Microinjections were attempted by three separate facilities, using two different methods. The Cornell Transgenic Core Facility did two rounds of standard microinjections and an additional round of microinjections using a laser-guided microinjection system. The Roswell Park Transgenic Core Facility did a single round of traditional microinjections, as well as a single round of laser-guided microinjections. Rebecca Holmes did two additional rounds of traditional microinjections. The injected blastocysts were implanted into pseudopregnant recipient mothers and allowed to come to

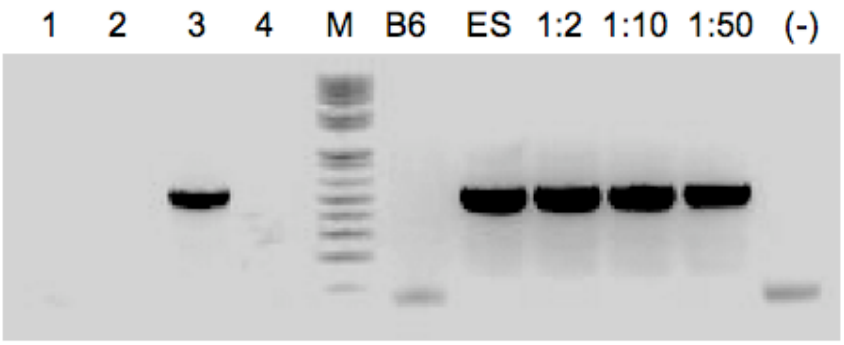


Figure III.2.5. PCR verification of a chimera containing the *Rasgrf1* DMD and repeats at the *Wnt1* locus. Lanes 1-4 show PCR amplification of 4 of the 19 live pups returned from microinjection using primers F1/FrtF, which are specific to the targeted allele. Animal number 3 (lane 3) is chimeric. Controls are to the right-hand side of the 1kb+ marker in lane 5. Controls include a wildtype B6 genomic DNA negative control (lane 6), targeted ES cells as a positive control (lane 7), 1:2 (lane 8), 1:10 (lane 9), and 1:50 (lane 10) mixtures of targeted ES cell DNA to genomic DNA as controls for PCR sensitivity, and a water negative control (lane 12).

term. Tail snips from all progeny were genotyped for chimerism using a PCR assay specific to the targeted allele. Of the 13 live and 11 dead pups recovered from the microinjections, 1 female was positive by tail snip PCR for the targeted allele (Figure III.2. 5). Nevertheless, all 13 live animals were bred to wildtype B6 mates and tail snips from all resulting pups were genotyped for inheritance of the targeted allele. None of the potential founders passed on a targeted allele to their progeny.

III.2.5 Discussion

Had this experiment been successful, the significance of this work would have been two fold. First, from a pure science standpoint, no other group has recapitulated imprinting at an endogenous locus in the genome. If the *Rasgrf1* DMD and repeats are sufficient to impart imprinting to a non-imprinted locus in the mouse genome, this would be the first example of such a success. The results would also confirm that no sequences in addition to the DMD and repeats are needed for proper imprinting and to solidify the current understanding of *Rasgrf1* imprinting control.

Unfortunately, no germline chimeras were obtained from multiple rounds of blastocyst microinjections. From 7 rounds of microinjection, only 19 live pups were obtained. This is dramatically lower than expected, as approximately 200 blastocysts were injected in total. However, there is a possibility that the presence of the *Neo* cassette, which operates under the control of its own promoter and enhancer, results in improper *Wnt1* expression. *Wnt1* is an important developmental signaling gene and ectopic

expression of *Wnt1* can enforce stem cell renewal and block differentiation, as can constitutive activation of beta catenin, a downstream target of *Wnt* signaling (Sato Nat Med 2004; Kielman 2002). Therefore, it is possible that ectopic expression of this gene could also lead to pre- or peri-natal lethality. Transfection of targeted ES cells with an expression vector containing the *Flpe* protein would effectively remove the *Neo* cassette from pCMM4.2 targeted ES cells. In future attempts to create targeted knock-in mice, ES cells containing a *Neo*-deleted version of the pCMM4.2 allele should be used in hopes of increasing the number of viable progeny. It is worth noting that a different restriction enzyme should be chosen for screening in this experiment, as the extra NdeI site used for previous Southern blot screening of targeted clones is contained within the Neo cassette.

One final concern that will persist even after deletion of the *Neo* cassette is that, like the transgene discussed above, the knock-in might faithfully recapitulate establishment of DNA methylation, but maintenance may fail. This would indicate that additional sequences outside of the DMD and repeats are needed for maintenance of DNA methylation.

III.3 LOCALIZATION OF A PIRNA WITH HOMOLOGY TO THE *RASGRF1* REPEATS

III.3.1 Abstract

Small RNAs are of interest in the field of epigenetics because of their roles in directing the placement of local epigenetic marks (Wasseneger 1994; Aufsatz 2002; Verdel 2004). One testis-specific category of small RNAs, piwi-associated RNAs (piRNAs), is suspected to play a role in the regulation of DNA methylation (Vagin 2006; Aravin 2008; Brennecke 2008). Using testis extracts and an LNA probe with homology to the *Rasgrf1* DNA repeats, a 31nt piRNA sized small RNA is detected by Northern blot. *In-situ* hybridization results indicate that this piRNA is expressed at P16 and P44, but not at P14, which coincides with the timing of expression of MIWI. These results suggest that piRNAs may be involved in epigenetic control of imprinted expression at *Rasgrf1*.

III.3.2 Introduction

A fundamental question in the field of epigenetics is how epigenetic marks are targeted to specific sequence locations. The *Rasgrf1* DNA repeats provide a unique tool to study the mechanisms by which epigenetic marks are regulated. These DNA repeats are necessary for establishment of DNA methylation in the male germline, are necessary for somatic maintenance of DNA methylation until implantation, are *cis*-acting, and control the epigenetic state of the surrounding region, up to several kilobase pairs away. As such, our lab began to search for proteins that associate with the repeats to identify

proteins that may be involved in the establishment and maintenance of DNA methylation at *Rasgrf1*.

The search for protein interactors was done in two ways. First, a variety of oligonucleotide probes corresponding to methylated and unmethylated versions of the repeat sequences were used in gel-shift assays with testes and brain extracts (Fitzpatrick and Soloway, unpublished). However, this method yielded nothing that bound either to only methylated or only unmethylated forms of the repeat probes. Candidates that did bind the repeat probes were readily competed away by non-specific probes. This approach was, therefore, abandoned due to specificity issues. The second approach used a yeast one-hybrid interaction trap. The DMD and the repeats were integrated at two reporter sites in the yeast genome and cDNA libraries from mouse testes, brain, and embryos to activate the reporters. This approach identified very few clones and each was identified only once. One clone did pass the dual reporter screen and, though the screen was for protein coding sequences, the sequence fused to the activation domain was not from a coding gene (Mahoney, Stablewski, Anggraini, and Drake, unpublished).

Since no repeat-binding proteins were identified other than CTCF, which had been shown previously to bind the repeats, the search turned instead to non-coding RNAs (Yoon 2005). Recently, noncoding RNAs have come to the forefront for their role in targeting chromatin remodeling complexes to imprinted genes regions. For example, at the *Xist* locus, the RepA repeats bind EZH2 within the polycomb repressive complex 2 (PRC2) (Zhao 2008). Likewise, at the *Slc22a1* locus, the *Air* noncoding RNA is

responsible for the recruitment of G9a to the *Slc22a1* promoter, and for the deposition of H3K9me3 (Nagano 2008). Additionally, small RNAs are known to regulate epigenetic phenomena in plants, *S. pombe*, and mammals (Wassenegger 1994; Aufsatz 2002; Verdel 2004; Aravin 2008; Carmell 2007), as well as to control physical interactions between distant regulatory elements in *Drosophila* (Brennecke 2007), making for a compelling case for a conserved mechanism across phylogenetic kingdoms.

LNA, or locked nucleic acid, probes contain modified RNA nucleotides. The ribose moiety of an LNA nucleotide is modified to contain an extra bridge linking the 2' and 4' carbons, which locks the ribose in the structural conformation often found in the A-form of DNA or RNA. The locked ribose conformation enhances base stacking and backbone pre-organization, significantly increasing the melting temperature of oligonucleotides containing LNA bases. Using LNA probes with homology to the *Rasgrf1* repeats, a conserved piRNA-sized 31nt RNA is detectable in testes extracts by Northern blot (Figure III.3.1 A-C; Lindroth, unpublished). Standard oligonucleotides of the same sequence gave no signal. Additionally, an RNA of the same size was identified using an LNA probe with homology to the *Igf2r* Region 2 ICR (Figure III.3.1 C; Lindroth, unpublished). However, these RNAs do not seem to eminate from the ICRs of these two loci, as mice homozygous for deletions of these ICRs still produce these RNAs (Lindroth, unpublished). piRNAs have been shown to control DNA methylation in mice in a MIWI-dependent manner. For example, mice lacking MIWI or MIWI2 show defects in methylation and silencing of L1 retrotransposons (Aravin 2007; Carmell 2007). In order to better understand the role of these small, testis specific piRNAs, and to

determine if their expression coincides with MIWI/MIWI2 expression, hybrid selection was carried out by Dr. Carlos Bosagna to try to clone and sequence the RNAs detected by the LNA probe, while I characterized the expression location of the *Rasgrf1* piRNA by *in situ* hybridization at several developmental time points.

III.3.3 Materials and methods

III.3.3.1 Tissue collection and sectioning

Testes were collected from FVB/n wildtype mice at P58, P14, and P16 Testes were snap frozen in liquid nitrogen and embedded in TissueTek O.C.T. Embedded tissues were transverse sectioned at 20um and at 8um and fixed using 4% paraformaldehyde in phosphate buffer (40mg/mL paraformaldehyde, 0.016M NaOH, 0.05M NaH_2PO_4, pH to 7.4 and filter sterilize).

III.3.3.2 In-situ hybridization

Fixed slides were washed three times in PBS and acetylated for 10 minutes at room temperarture (270mL DEPC H20, 30mL TEA pH 7.8, add 750uL 95% acetic anhydride and mix well just before application to slides). Slides were washed twice in PBS and dried. A 3' biotin-labeled LNA probe (PDS224 5' – CAG cCG cTA cTG cTG cCC cTG cCC cTC - 3', where nucleotides in lowercase are LNA modified) was prepared by denaturing at 80C in hybridization buffer (0.1% Tween 20, 50% formamide, 5X SSC, %X Denhardts, 5mM EDTA, 10mM pH 8.0 NaH_2PO_4, 250ug/mL salmon sperm DNA, 100ug.mL tRNA, 100ug/mL yeast RNA) then cooling on ice and was

hybridized to the sections overnight at 55C in hybridiztion buffer. Slides were incubated in a humidified chamber. Slides were washed in 0.2X SSC**,** followed by B1 (0.1M Tris, 0.15M NaCl). Slides were blocked for 1hr with blocking solution (10ml B1 plus 0.2g of Roche blocking reagent, #1 096 176, and 0.1mL HINGS) and 120ul of a 1:3000 dilution of anti-digoxygenin-AP Fab fragments (Roche # 1 093 274) to blocking solution was added to slides overnight at 4C. Slides were stained with 100ul of staining solution (1ml B3 buffer (0.1M Tris, 0.1M NaCl, 0.05M MgCl2), 14ul NBT (Promega, #S380, 50mg/mL), 7.2ul BCIP (Promega #S381C, 50mg/mL) in the dark. Slides were washed in PBS without MgCl2, re-fixed in 4% paraformaldehyde overnight, and washed in PBS again. Coverslips were mounted using 55ul 70% glycerol and nail polish (Wet and Wild clear does not auto-fluoresce).

III.3.3.3 Slide staining

Slides were stained with hematoxylin and eosin to visualize nuclei and cytoplasm. Slides were immersed in filtered Harris hematoxylin for 1min, rinsed with tap water, and immersed in 1% eosin (1g eosin dye, 100ml deionized H2O) for 1-2 minutes. Slides were rinsed with tap water and dehydrated through 50%, 70%, 80%, two changes of 95%, and two changes of 100% ethanol solutions, followed by two rinses in xylenes.

Slides were stained with acridine orange to visualize nucleic acids. Slides were immersed in acridine orange dye (0.05g acridine orange dye, 5mls glacial acetic acid, 500mls H2O) for 30min and rinsed in 0.5% acetic acid in 100% alcohol, followed by two rinses in 100% ethanol and two rinses in xylenes.

Slides were stained with methyl green to visualize nuclei. Slides were immersed in distilled water followed by methyl green solution (0.5g ethyl violet free methyl green, 1.36g sodium acetate, 100ml distilled H2O pH'ed to 4.2) for 5 minutes. Slides were rinsed in distilled water and dehydrated through 95% ethanol followed by two changes of 100% ethanol and cleared in xylenes.

III.3.4 Results

Acridine orange staining indicated that there was no RNA degradation during sectioning or fixation of the tissue sections to the slides (Figure III.3.2 A, D, G). The results of the *in-situ* hybridization with the LNA repeat probe indicate that the putative *Rasgrf1* piRNA is expressed in testis, confirming the Northern blot results performed by Dr. Anders Lindroth (Figure III.3.1 B). Based on the cell morphology observed in the hematoxylin and eosin staining and the methyl green staining, expression appears to be cytoplasmic and expression is observed at P16 and P44, but not at P14 (Figure III.3.1 B-C, E-F, and H-1 and III.3.3 A-C). Controls done with probe but without anti-digoxygenin-AP Fab antibody and without probe but with anti-digoxygenin-AP Fab antibody yielded no staining, indicating that staining is due to hybridization of the probe, and not due to artifacts (Figure III.3.4 A-B).

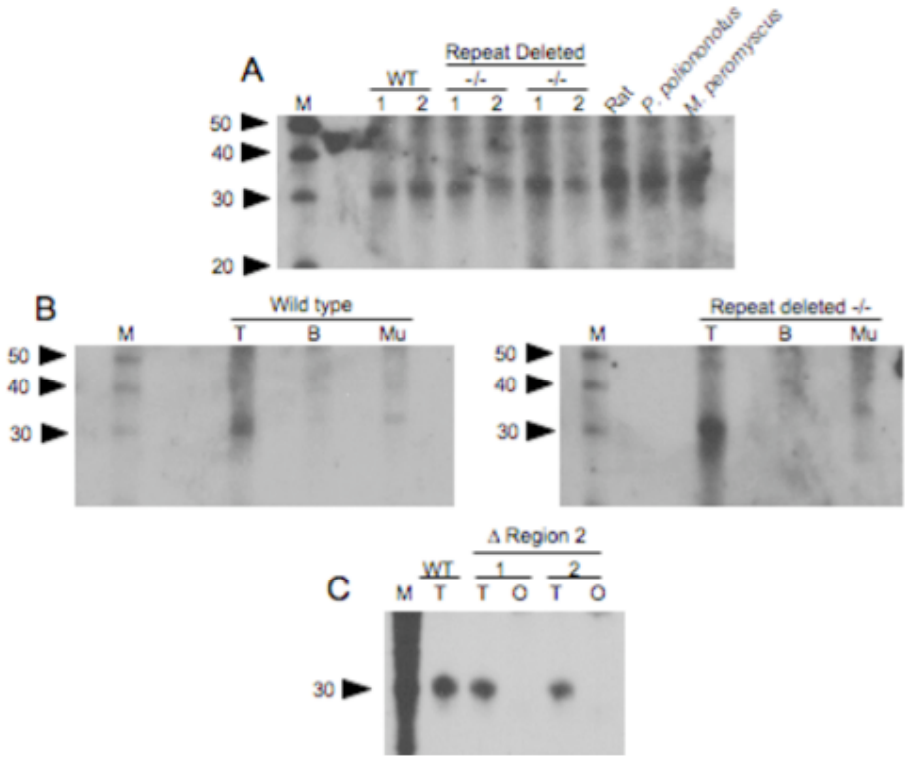


Figure III.3.1. A conserved piRNA with homology to the *Rasgrf1* DNA repeats is expressed in the testis in both wildtype and repeat deleted animals[3]. (A) Northern blot analysis using a probe with homology to the *Rasgrf1* repeats reveals that a small, 31nt RNA is detected in testis of *M. musculus*, rat, *P. polionotus* and *M. peromyscus.* Additionally, the 31nt RNA is expressed in the testis of mice homozygous for a deletion of the *Rasgrf1* repeats. (B) Northern blot analysis comparing expression profiles of the 31nt RNA in wildtype and mice homozygous for a deletion of the *Rasgrf1* repeats. The 31nt RNA is strongly expressed in testis in both genotypes. Expression in muscle is very low and expression in brain is not detected. (C) Northern blot analysis using a probe with homology to the *Igf2r* Region 2 ICR. As with the 31nt RNA with homology to *Rasgrf1*, the 31nt RNA with homology to Region 2 is expressed in both wildtype mice and mice homozygous for a deletion of Region 2.

[3] Northern blots by Dr. Anders Lindroth.

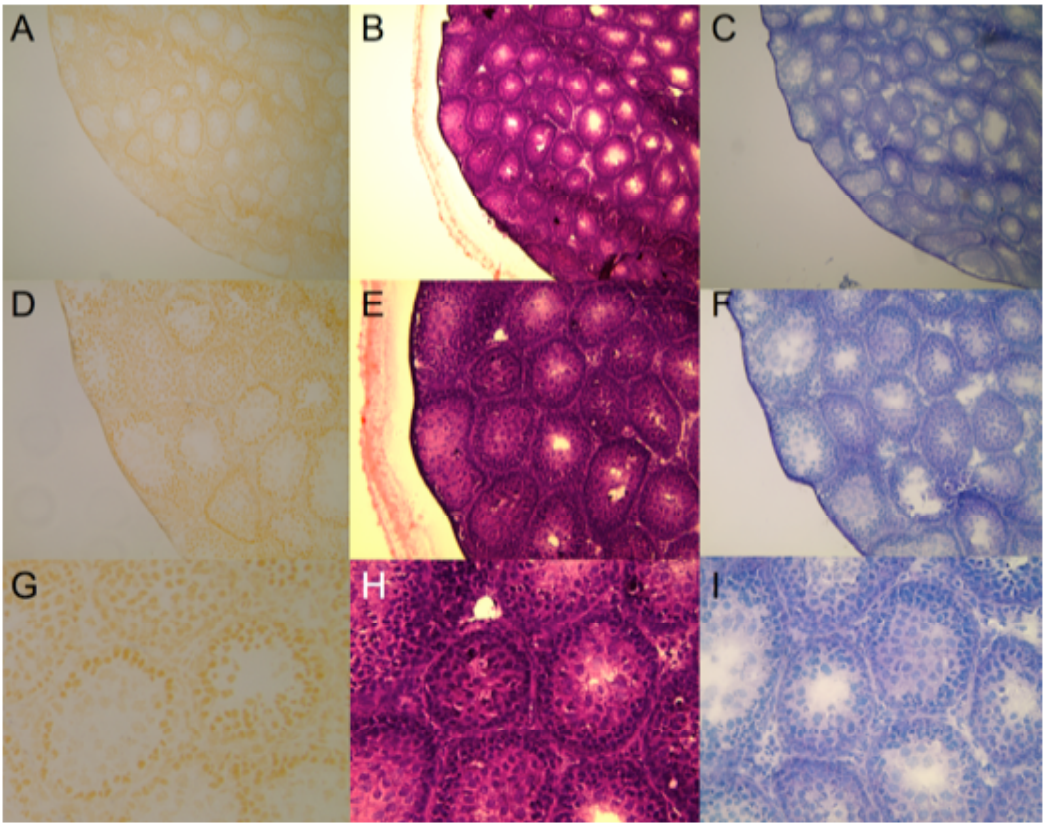


Figure III.3.2. Acridine orange, H&E, and methyl green staining of testis sections. (A, D, G) Representative acridine orange staining of 8um testis sections at 10X (A), 20X (D), and 40X (G). Acridine orange stain intercalates into double-stranded nucleic acid. When viewed under fluorescence, acridine orange will appear green when bound to DNA and orange when bound to RNA (not shown). (B, E, H) Representative hematoxylin and eosin (H&E) staining of 8um testis sections at 10X (B), 20X (E), and 40X (H). Hematoxylin colors basophilic structures blue, and alcohol-based acidic eosin Y colors eosinophilic structures pink. Basophilic structures usually contain nucleic acids and the cytoplasmatic regions are usually rich specifically in RNA. (C, F, I) Representative methyl green staining of 8um testis sections at 10X (C), 20X (F), and 40X (I). Methyl green stains nuclei a light blue-green color. All sections shown are from an adult P58 wildtype male.

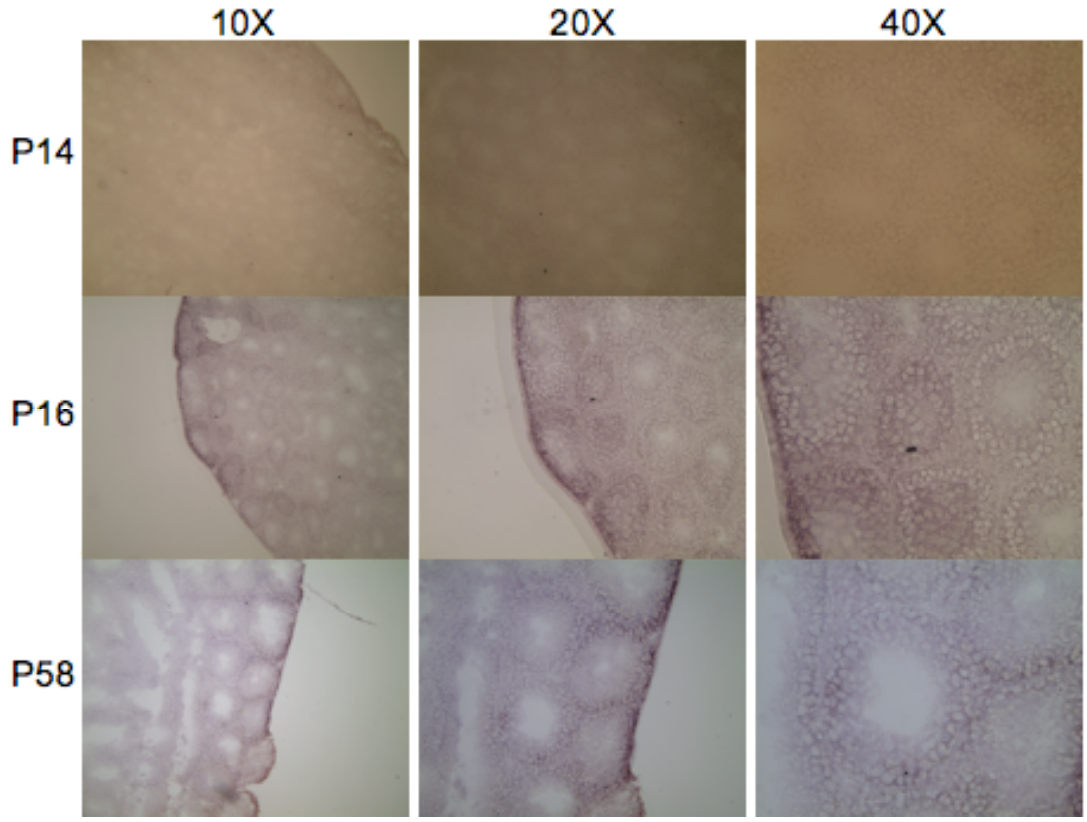


Figure III.3.3. *In-situ* hybridization timecourse for expression and localization of a 31nt piRNA with homology to the *Rasgrf1* DNA repeats. *In-situ* hybridization was done using an LNA probe with homology to the DNA repeats. Sections are 8um and were done for three developmental time points: P14, P16, and P58. Each developmental time point is accompanied by images at 10X, 20X, and 40X magnification. The *Rasgrf1* piRNA appears to be cytoplasmic and is not expressed until P14. This developmental timing of expression corresponds to the expression pattern of MIWI, which is involved in the production of piRNAs (Kuramochi-Miyagawa 2001).

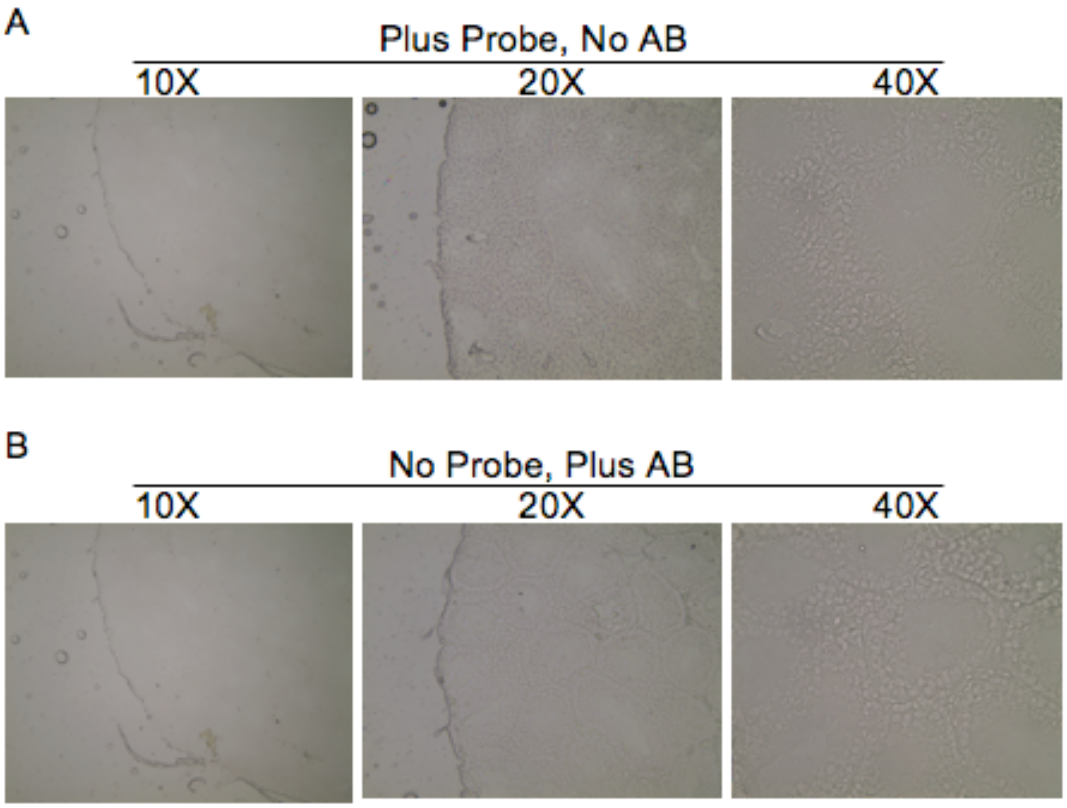


Figure III.3.4. *In-situ* control staining lacking either probe or antibody. (A) Representative *in-situ* staining done using an LNA probe with homology to the *Rasgrf1* DNA repeats, but lacking anti-digoxygenin antibody, at 10X, 20X, and 40X. (B) Representative *in-situ* staining done with an anti-digoxygenin antibody, but lacking an LNA probe with homology to the *Rasgrf1* DNA repeats, at 10X, 20X, and 40X. No coloration is visible in either control, indicating that the staining seen in Figure III.3.3 is not an artifact. All sections shown are from an adult P58 wildtype male.

III.3.5 Discussion

Using testis extracts and an LNA probe with homology to the *Rasgrf1* DNA repeats, a 31nt piRNA sized small RNA is detected by Northern blot (Lindroth, unpublished). *In-situ* hybridization results indicate that this piRNA is expressed at P16 and P44, but not at P14, which coincides with the timing of expression of MIWI and lends support to the identification of this small RNA as a piRNA (Kuramochi-Miyagawa 2001). Interestingly, there is a link between MIWI and DNA methylation, as mice lacking MIWI or MIWI2 similarly show defects in methylation and silencing of L1 retrotransposons (Aravin 2007; Carmell 2007). As imprinted genes depend heavily upon allele-specific DNA methylation to mediate allele-apecific expression, it would be interesting to test whether knockdown of MIWI abolishes expression of the *Rasgrf1* piRNA.

Another unanswered question is from where does the *Rasgrf1* piRNA emanate? Deletion of either the *Rasgrf1* or the *Igf2r* ICR does not abolish expression of the piRNAs with homology to these regions, indicating that expression is not from the ICRs. The *Xist* repeat B unit shares homology with the *Rasgrf1* repeats, so it is possible that the piRNA is expressed from this locus, or another locus with homology to the *Ragsrf1* repeats, and is targeted to the *Rasgrf1* locus based on sequence homology (Lindroth, unpublished). There is evidence for this mode of action in *Drosophila.* A piRNA cluster on the X chromosome in *Drosophila* can regulate the activity of P-elements throught the genome (Brennecke 2007). One way to address this possibility is by cloning and sequencing the *Rasgrf1* piRNA, using existing methods (Ro 2007). This was attempted once by Dr. Carlos Bosagna, but all RNAs identified by this method had no more similarity to the *Rasgrf1* repeats, based

on ClustalW alignments, than did randomly generated DNA sequences. This raises the issue of whether the *Rasgrf1* LNA probe lacks specificity and detects multiple piRNAs, which may also be represented in the *in-situ* expression profiling. In either case, it would be worth attempting the hybrid selection again. Once the locus of origin for this piRNA is identified, knockout studies can be done to determine whether it affects DNA methylation at *Rasgrf1.*

III.4 NECESSITY OF DNA REPEATS FOR A *TRANS*-EXPRESSION EFFECT AT *AK029869*

This project was started by a former postdoc in the lab, Krista Kauppinen, and I took over after she left. Krista discovered that *AK029869* is imprinted and subject to a *trans*-expression effect. I confirmed the imprinted status of *AK029869* as well as one of the *trans*-effects. Krista also performed 3C analysis on the *AK029869* imprinted region and I began to do FISH analysis to complement those results. I contributed Figure III.4.1, III.4.2, Figure III.4.10 and III.4.11 and I confirmed data in Figure III.4.3, Figure III.4.5 and III.4.7.

III.4.1 Abstract

Imprinted genes are monoallelically expressed, with expression patterns dependent on parent-of-origin. Often, imprinted genes are located in clusters near imprinting control centers (ICR) that control imprinted expression along the entire cluster. One such cluster, consisting of *A19* (also called *AK015891*) and *Rasgrf1* is located on mouse chromosome 9. *A19* is a paternally expressed noncoding RNA located approximately 46kb upstream of *Rasgrf1*, a paternally expressed Ras GTPase. In this genomic region exist four additional noncoding RNAs (ncRNAs), one of which, *AK029869*, is imprinted and paternally expressed. Intriguingly, any of several maternally inherited mutations to the wildtype ICR silence paternal expression of *AK029869* in *trans.* This constitutes the second observed *trans* expression effect in this region and I aimed to explore when during early development the *Rasgrf1* DNA repeats are needed in order to prevent improper *trans*-

Figure III.4.1. Schematic of the *Rasgrf1* imprinted region. Screenshot of the UCSC Genome Browser depiction of the Rasgrf1 imprinted region. *Rasgrf1* is an imprinted and paternally expressed gene on mouse chromosome 9. The *Rasgrf1* ICR is located approximately 30kb upstream of the *Rasgrf1* coding region. Upstream of the ICR are four non-coding RNAs, two of which are imprinted and paternally expressed: *AK029869* and *AK015891*. All three imprinted transcripts are controlled by the *Rasgrf1* ICR, which consists of a DMD and a series of 40 copies of a 41bp repeat unit.

expression effects, and whether the DNA repeats mediate physical interaction between the maternal and paternal alleles.

III.4.2 Introduction

The *Rasgrf1* ICR is located approximately 30kb upstream of the *Rasgrf1* transcription start site, and lies between *Rasgrf1* and five upstream ncRNAs (Figure III.4.1). The ICR contains a differentially methylated domain (DMD) as well as a series of tandem DNA repeats, consisting of 40 copies of a 41bp repeat unit (Yoon 2002). The DMD acquires DNA methylation exclusively on the paternal allele and placement of DNA methylation is controlled by the tandem repeats (Figure III.4.2; Yoon 2002). The repeats are required for both establishment and maintenance of allele-specific DNA methylation during spermatogenesis and through pre-implantation development and constitute the only sequence thus far that is known to be necessary for the placement of DNA methylation along a DMD (Holmes 2006; Yoon 2002). Imprinted expression is a result of both allele-specific binding of the methylation-sensitive enhancer blocking protein CTCF and allele-specific DNA methylation. CTCF binds the unmethylated maternal allele DMD and disrupts enhancer to promoter interaction, resulting in silencing of *Rasgrf1*. On the paternal allele, the DMD is methylated, preventing binding of CTCF and allowing paternal allele expression (Yoon 2005). The presence of the DNA repeats on the paternal allele is necessary for the establishment and for the maintenance of DNA methylation at the DMD until approximately e5.5. Deletion of the paternal repeats before this time point leads to loss of DNA methylation on the paternal allele, resulting in silencing of *Rasgrf1* (Holmes 2006). The DNA repeats, in

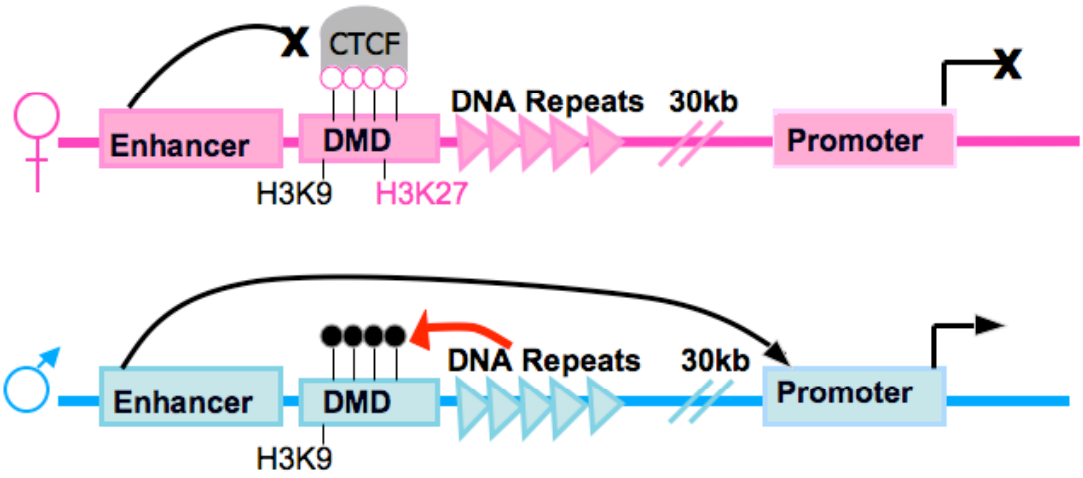


Figure III.4.2. Current model for epigenetic control of imprinted expression at *Rasgrf1.* One the paternal allele (depicted in blue), a series of DNA repeats are responsible for the establishment and maintenance of paternal allele-specific DNA methylation at CpG dinucleotides in the DMD (Yoon 2002; Holmes 2006). Placement of DNA methylation at the DMD prevents CTCF binding and allowing promoter-enhancer interaction, resulting in transcriptional activation (Yoon 2005). Placement of DNA methylation at the DMD also blocks the binding of H3K27me3 (Lindroth 2008). The maternal allele (depicted in pink) is unmethylated at CpG dinucleotides in the DMD. The unmethylated status if the DMD allows binding of CTCF and prevents the *Rasgrf1* upstream enhancer from interacting with the promoter, resulting in suppression of transcription (Yoon 2005). Lack of methylation at the DMD also allows placement of the silencing mark H3K27me3 (Lindroth 2008). H3K9me3 is found on both the maternal and the paternal allele and therefore does not appear to control differential expression. Circles above the DMD represent the methylation status of CpG dinucleotides within the DMD. Black circles represent a methylated DMD and white circles represent an unmethylated DMD.

combination with the DMD constitute a binary switch that is necessary for imprinted expression of *Rasgrf1*.

Our lab has shown previously that replacing the *Rasgrf1* repeats with the Region 2 differentially methylated region (Region 2 allele) from the imprinted and maternally expressed gene, *Igf2r*, allows for DNA methylation and expression of the paternal allele of *Rasgrf1* (Herman 2003). Paternal transmission of the Region 2 allele also results in both methylation and expression, in *trans,* of the normally unmethylated and silent wildtype maternal allele. Additionally, the activated wildtype maternal allele maintains its active state in subsequent generations, even in the absence of a Region 2 paternal allele. These results replicate several features common to a phenomenon described as paramutation in plants and constitute the first example of a *trans*-expression effect in this genomic region.

In addition to *Rasgrf1* is a second known imprinted gene in this region, *A19* (also called *AK015891*), a ncRNA located ~10 kb upstream of the ICR. *A19* is highly expressed in testis and is also expressed in neonatal and adult brain, although at lower levels (de la Puente 2002). Imprinting of *A19* occurs in brain, where expression is solely from the paternal allele (de la Puente 2002). Our lab has identified a second ncRNA, *AK029869,* located ~5kb upstream of the *Rasgrf1* ICR that is also imprinted and paternally expressed in neonatal brain. Furthermore, we have shown that several maternal allele ICR mutations lead to silencing of the normally expressed *AK029869* paternal allele in *trans* and preliminary data suggest the possibility that the *trans* expression effect is dependent on physical interaction between homologues.

III.4.3 Materials and methods

III.4.3.1 SNP and restriction site identification

SNPs were identified in *AK029869* using the Jackson Laboratory Mouse Genome Informatics website (http://www.informatics.jax.org). SNPs available between C57/BL6 and PWK mouse strains were noted. In each case, potential SNPs were analyzed using NEB cutter (http://tools.neb.com/NEBcutter2/index.php) to select SNPs overlapping a restriction enzyme recognition site for identification of allele-specific expression.

III.4.3.2 Tissue collection

Crosses were set up between either wildtype B6 abd PWK mice or B6 mice homozygous for loxP-flanked copies of the *Rasgrf1* DNA repeats and PWK mice carrying either *Zp3 Cre* or *Meox2 Cre* alleles. In each case, crosses were set up as reciprocal pairs to either rule out expression differences due to strain QTLs or to examine the effect of inheritance of both maternal and paternal repeat deletions. The progeny of each of these crosses were sacrificed at P10 (except for the wildtype imprinting timecourse experiment) and a small portion of the brain was collected for genotyping. The remainder of the brain was snap frozen in liquid nitrogen for later use.

III.4.3.3 Genotyping

Brain DNA samples were prepared by lysing in Larid's lysis buffer plus proteinase K overnight followed by ethanol precipitation. Brain DNA was

genotyped for the presence of one B6 allele and one PWK allele using primers either AKnewFWD (5'- CTT TCT CAA GCA ACC TAT C -3') and AKnewREV (5'- AAG GAC CTG CCG CTT AAC T -3') or primers PDS155 (5'- ATT CAC CGC TGC TGC TTA AA -3') and AKR1-KPK (5'- TAG GAA AAT GGC TCG GTG TC -3') for 40 cycles under the conditions 94C for 30sec, 60C for 1 min, 72C for 2min. Also, for the repeat-deletion experiments, deletion of the DNA repeats was determined using the primer combination PDS16 (5' – GCA CTT CGC TAC CGT TTC GC – 3'), PDS18 (5' – TTT CTG CCA TCA TCC CAG CC – 3'), and PDS17 (5' – TGT CCT CCA CCC CTC CAC C– 3') and cycling conditions 94C for 10sec, 61C for 20sec, 72C for 50sec for 40 cycles.

III.4.3.4 RNA preparation

Brain samples were isolated from F1 progeny of reciprocal crosses at P10 (except in the case of the imprinting timecourse experiment) and total RNA was prepared. For each neonatal brain, 2mls of GTC RNA lysis buffer was used (4M guanidium thiocyanate, 25mM pH 7.0 sodium citrate, 100mM beta-mercaptoethanol, 0.5% sarcosyl, 0.2M pH 4 sodium acetate, and 50% acidic phenol) and each brain was homogenized for 45 seconds at 18,000rpm. Following homogenization, RNA was extracted with 0.2 volumes chloroform followed by isopropanol precipitaion. RNA was resuspended in 10mM Tris-EDTA.

III.4.3.5 cDNA analysis

cDNA was prepared from 5ug of RNA treated with 2.5ul of DNaseI (Invitrogen). Amplification was done using random primers (Invitrogen) and

Superscript II reverse transcriptase (Invitrogen). Following cDNA synthesis, nested PCR was performed using 0.5ul cDNA as template. First round PCR was done with primers PDS155 (5' - ATT CAC CGC TGC TGC TTA AA – 3') and AKR1-KPK (5' – TAG GAA AAT GGC TCG GTG TC – 3') for 19 cycles. 2 ul of first round PCR product was diluted into 18ul of water, and 1.5ul of this dilution was used as template for 35 cycles of second round PCR. Second round PCR primers were AKnewFWD (5' – CTT TCT CCA GCA ACC TAT C) and AKnewREV (5' – AAG GAC CTG CCG CTT AAC T – 3'). In each case, cycling conditions were 94°C, 60°C, 72°C / 30 seconds, 1 minute, 2 minutes. 10ul of second round PCR product was digested 5hours to overnight with 1U AluI (NEB). Digests were heat inactivated and run on a 3% agarose gel.

III.4.3.6 3C analysis

Nuclei were isolated from a wildtype P10 B6/PWK F1 mouse brain. Briefly, cells were fixed with 2% formaldehyde at room temperature for 5 minutes, quenched with 0.125M glycine, and pelleted at 3500rpm for 15 minutes. After pelleting, the cells were lysed on ice for 90 minutes in 50mls lysis buffer (10mM pH 8.0 Tris-HCL, 10mM NaCl, 0.2% NP-40, 0.1mM PMSF, 1:500 Sigma protease inhibitor cocktail). After lysis, nuclei were collected by spinning for 15 minutes at 2500rpm. Nuclei were incubated in 1ml of Neb Buffer 3 plus 0.3% SDS at 37C for one hour with shaking. 1.8% Triton-X was added and the reaction was incubated at 37C one hour with shaking. Nuclei were counted with a haemocytometer and $1x10^8$ nuclei (~15ug) were used for restriction enzyme digestion. Digestion was done overnight at 37C with shaking. To inactivate the restriction enzyme, 1.6% SDS was added and the reaction was heated to 80C for 20 minutes. 20ul (2ug) of chromatin was used

with 80ul of 10X T4 ligase buffer and 694ul DEPC H2O to make up an 800ul ligation reaction. 1% Triton-X was added and the reaction was incubated at 37C for one hour. The temperature was lowered to 16C and 6ul T4 ligase was added and incubated for four hours. The ligated DNA was purified by incubating with 100ug/mL proteinase K and incubating at 65C overnight. The sample was treated with 0.5ug/mL RNase A for 30minutes at 37C. DNA was extracted by phenol:chloroform followed by ethanol precipitation. Captured DNA was subjected to PCR using various combinations of forty primer pairs. PCR was done under the cycling conditions 94C for 30 seconds, 61C for one minute, 72C for two minutes for forty cycles. Controls included samples that had undergone ligation but not crosslinking, and samples that had undergone crosslinking but not ligation.

III.4.3.7 *Rasgrf1* ICR FISH

2ug of BAC RP24-228H14 was labeled with both Spectrum green-dUTP (Vysis, 30-803200) and Spectrum orange-dUTP (Vysis, 30-803000) using the Vysis FISH nick translation kit (Vysis, 32-801300). Labeled probe was pre-hybridized with 20ul CotI DNA (Gibco, 18440-016) and precipitated with 3M NaOAc and 100% ethanol. 40ul of hybridization buffer (0.5ml formamide, 0.1ml 20X SSC [175.32g NaCl, 88.23g sodium citrate, DEPC water to 800mls, pH to 7.0 and bring volume to 1L with DEPC water], 0.25ml 40% dextran sulfate-500K, 0.1ml of 20mg/ml BSA, 0.05ml DEPC water, store at -20C) was added and the probe was heated to 42C for 10 min to resuspend, and aliquoted into 6ul batches stored in the dark at -20C.

MEFs were grown in T75 flasks until confluent and were trypsinized,

counted, and resuspended at 1x10^3 cells/ul. 125ul was cytospun onto each of eight slides. Slides were fixed by immersing in ice cold autoclaved PBS (8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4, adjust pH to 7.4, bring volume to 1L with DEPC water, autoclave) for 5 minutes, ice cold CSK buffer (5.84g NaCl, 102.7g sucrose, 3.02g PIPES, 0.61g $MgCl_2 \cdot 6H_20$, adjust pH to 6.8, bring final volume to 1L with DEPC water and autoclave, store at 4C) for 1 minute, ice cold CSK plus Triton-X (5.84g NaCl, 102.7g sucrose, 3.02g PIPES, 0.61g $MgCl_2 \cdot 6H_20$, 5mls Triton-X, adjust pH to 6.8, bring final volume to 1L with DEPC water and autoclave, store at 4C) 1 minute, and ice cold CSK 1 minute. Slides were fixed for 10 minutes in 4% paraformaldehyde (2g paraformaldehyde, 50mls PBS, pH to 7.4 and filter sterilize) and washed in 70% ethanol for 5-10 minutes. Slides can be stored under 70% ethanol at 4C at this point. Slides were dehydrated through 80%, 90%, and 100% ethanol at room temerature for 2 minutes each. Slides were denatured in 70% formamide (75ml formamide, 5ml 20XSSC [175.32g NaCl, 88.23g sodium citrate, DEPC water to 8oomls, pH to 7.0 and bring volume to 1L with DEPC water], 10ml water) for 10 minutes at 80C. Denaturation was quenched with an ice cold ethanol series of 70%, 80%, 100%, and 100% for two minutes each. Probe was diluted in 1:5 in DEPC water to create a working stock and heated to 80C for 10 minutes, followed by incubation at 37C until ready to use. Slides were air-dried and 10ul of dilute probe was added to each slide, coverslipped, and sealed with rubber cement. Slides were placed in an empty pipette tip box with water added under the tip platform and incubated at 37C overnight.

The next day, rubber cement was removed and slides were dunked in

50% formamide (75ml formamide, 60ml DEPC water, 15ml 20X SSC) to remove coverslips. Slides were washed three times in 50% formamide at 45C for 5 minutes each. Then, slides were washed in fresh 2X SSC (10mls 20X SSC, 90mls DEPC water) at 45C for 5 minutes each. Slides were air-dried, DAPI plus antifade was added (Vectashield, H-1200) and slides were coverslipped and sealed with Wet 'N Wild clear nailpolish. Slides were allowed to sit at room temperature before viewing or storing in the dark at -20C.

III.4.4 Results

III.4.4.1 SNP and restriction site identification

To test the imprinting status of *AK029869*, we first identified useful SNPs between the polymorphic mouse strains PWK and B6. After identifying possible SNPs, one was chosen that overlapped with an allele-specific restriction enzyme site. *AK029869* is an unspliced transcript, so primers could be tested on genomic DNA before use on cDNA for expression analysis. In these initial primer tests, we included an analysis for strain-specific amplification bias. This analysis consisted of mock F1 genomic DNA made from mixing genomic DNA from the two pure inbred parental strains at 3:1 and 1:3 ratios and demonstrated a PWK specific strain amplification bias (Figure III.4.3).

III.4.4.2 Allele-specific expression of *AK029869*

Reciprocal crosses between polymorphic mouse strains were used and

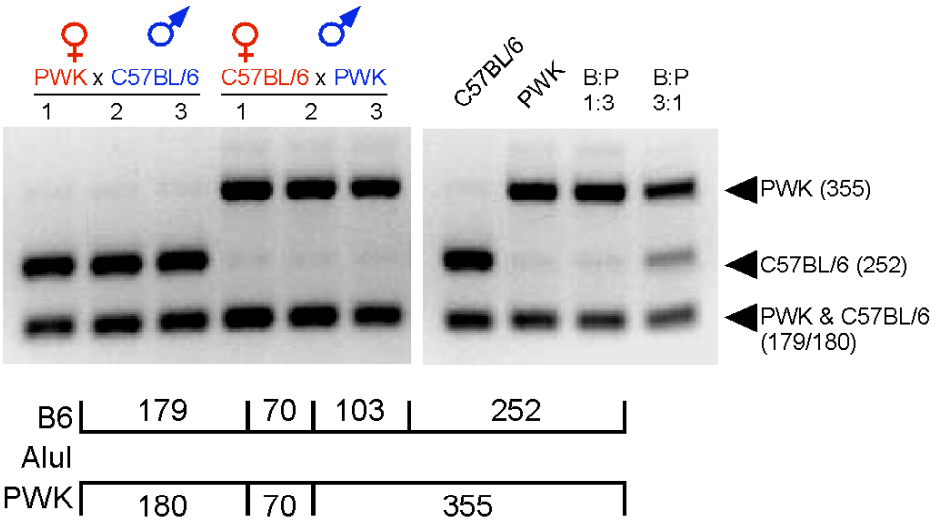


Figure III.4.3. *AK029869* is imprinted and shows a slight PWK strain-specific amplification bias[4]. Brains from P10 progeny of B6 pure inbred (lane 1, right hand gel), PWK pure inbred (lane 2, right hand gel), B6 x PWK F1 (lanes 1-3, left hand gel) and PWK x B6 F1 (lanes 4-6, left hand gel) animals were dissected out and RNA and cDNA was prepared. Nested RT-PCR was performed on the cDNAs and the resulting PCR products were digested with AluI. Expression from the B6 allele results in the following banding pattern: 252bp, 179bp, 103bp, 70bp (lane 1, right hand gel). Expression from the PWK allele results in the following banding pattern: 355bp, 180bp, 70bp (lane 2, right hand gel). Expression from the B6 x PWK cross shown in lanes 1-3 of the left hand gel is consistent with paternal PWK expression. Expression from the reciprocal PWK x B6 cross shown in lanes 4-6 of the left hand gel is consistent with paternal B6 expression. Lane 3 of the right hand gel shows amplification of a mock F1 consisting of a 1:3 mixture of B6:PWK genomic DNA. Lane 4 of the right hand gel shows amplification of a mock F1 consisting of a 1:3 mixture of PWK:B6 genomic DNA. The mock F1 analysis indicates that there is a slight amplification bias of the PWK allele. However, the bias is not severe enough to confound imprinted expression assays.

[4] Figure from Dr. Krista Kauppinen.

brains of neonatal progeny were collected for analysis. As imprinted expression can be developmental time point specific, we collected tissue samples beginning at e16.5 and continuing through P42. After cDNA synthesis, the templates were subjected to nested PCR followed by digestion with an allele-specific restriction enzyme, *AluI*. Beginning at e16.5 and continuing through P42, *AK029869* demonstrated paternal allele specific expression (Figure III.4.4).

III.4.4.3 Allele-specific expression of *AK029869* depends on paternal DNA repeats

Next, we wanted to test whether paternal-allele specific expression of *AK029869* depends upon the presence of the tandem DNA repeats in the ICR, similar to *Rasgrf1*. To do this, we used mice carrying a targeted deletion of the tandem DNA repeats (repeat-deleted allele) in reciprocal crosses to the polymorphic PWK mouse strain carrying the allele-specific SNP mentioned above (Figure III.4.3 and Figure III.4.5). As with *Rasgrf1*, paternal inheritance of a repeat-deleted allele leads to silencing, in *cis*, of *AK029869*. Expression of the normally silent maternal allele is unaffected, as is DNA methylation at the maternal DMD. Unexpectedly, maternal inheritance of the repeat-deleted allele leads to silencing in *trans* of AK029869 (Figure III.4.6 and Figure III.4.7). This *trans*-silencing seems to be independent of DNA methylation, as maternal inheritance of a repeat deleted allele does not change the methylation status of the DMD (although it affects H3K27me3, discussed in Appendix Chapter III.1 above), but does lead to silencing of the paternal allele (Figure III.4.7).

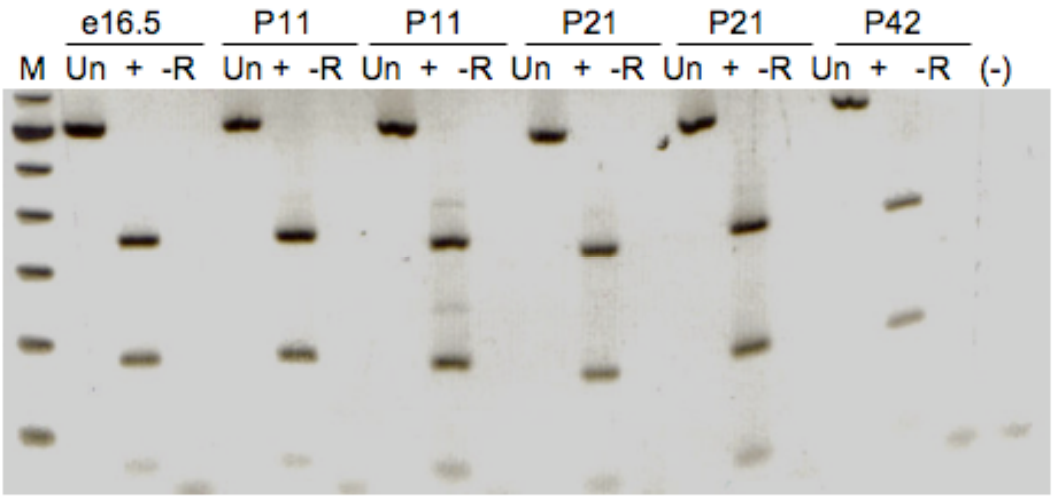


Figure III.4.4. Developmental timecourse demonstrating timing of imprinted AK expression[5]. Brains from a B6 x PWK cross were dissected out at e16.5 (+, lane 3), P11(+, lanes 6 and 9), P21 (+, lanes 12 and 15), and P42 (+, lane 18). RNA and cDNA were prepared from brain samples and cDNA was subjected to nested RT-PCR. The resulting PCR products were digested with AluI. All samples from e16.5 to P42 show imprinted and paternal allele-specific expression, indicated by the elclusively PWK banding pattern of 334bp, 180bp, and 70bp. Also shown for each sample is the undigested PCR product (Un, lanes 2, 5, 8, 11, 14, and 17) and amplification of the –RT sample (-R, lanes 4, 7, 10, 13, 16, and 19).

[5] Figure from Dr. Krista Kauppinen.

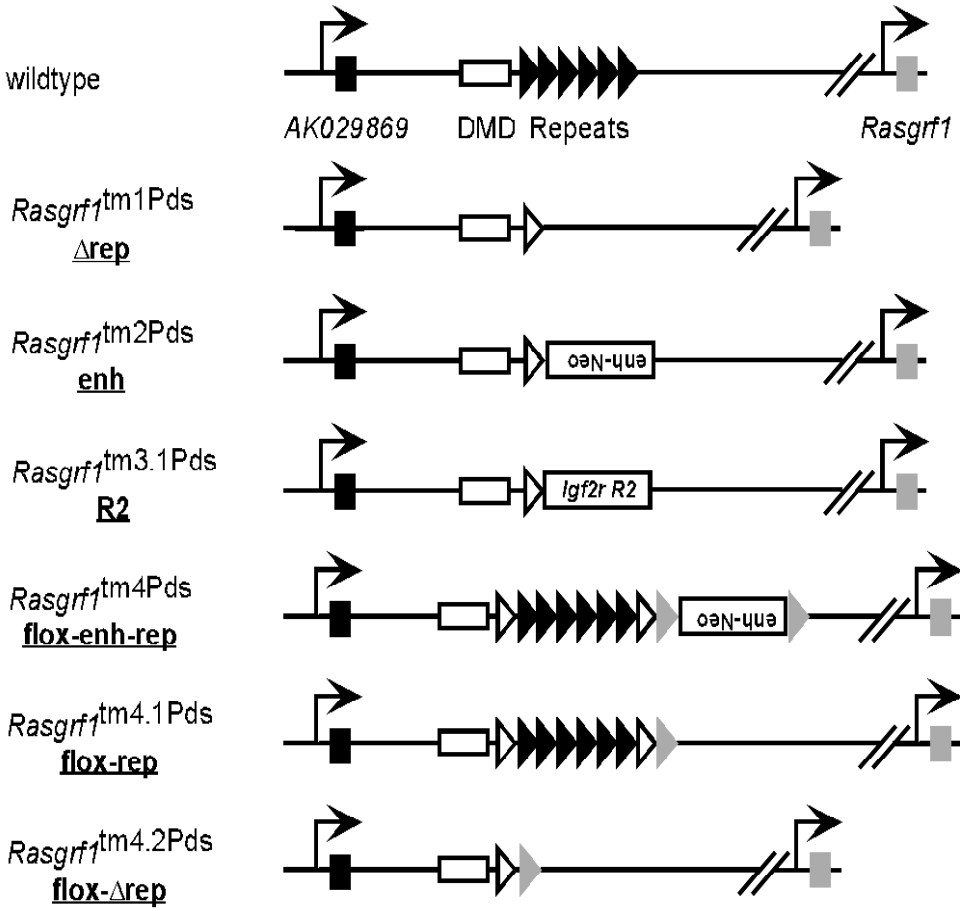


Figure III.4.5. Wildtype and mutant alleles of the *Rasgrf1* ICR[6]. Schematic showing wildtype and the various *Rasgrf1* ICR mutant alleles. The upstream *AK029868* transcript (black rectangle) and the downstream *Rasgrf1* transcript (grey rectangle) are shown, as are the DMD (white rectangle) and the repeats (tandem series of black triangles). Arrows directly to the left of the transcripts indicate the direction of transcription. Allele names are to the left of each of the cartoon alleles.

[6] Figure from Dr. Paul Soloway.

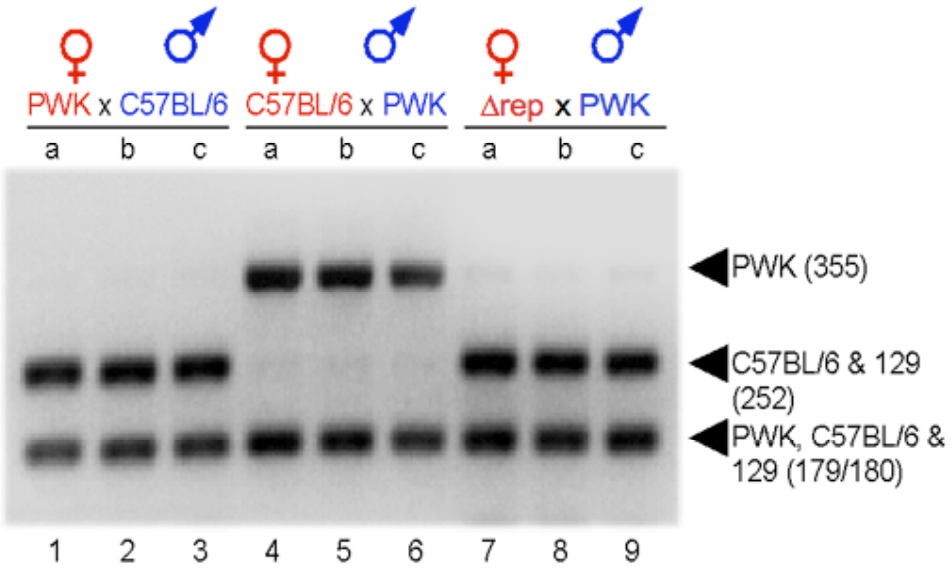


Figure III.4.6. Paternal expression of *AK029869* is silenced in *trans* by a maternal allele mutation[7]. Brains of progeny from PWK x B6, B6 x PWK, and Δrep x PWK crosses were dissected out at P10. RNA and cDNA was prepared from brain tissue and nested RT-PCR was performed on the cDNA. Resulting PCR products were digested with AluI. Samples from the PWK x B6 cross (lanes 1-3) show B6 expression, consistent with imprinted paternal expression. Samples from the reciprocal B6 x PWK cross (lanes 4-6) show B6 expression, consistent with imprinted paternal expression. Progeny from the Δrep x PWK cross show B6 expression, consistent with silencing of the normally expressed paternal allele and expression of the normally silenced maternal allele. The Δrep allele is represented in cartoon format in Figure III.4.5.

[7] Figure from Dr. Krista Kauppinen.

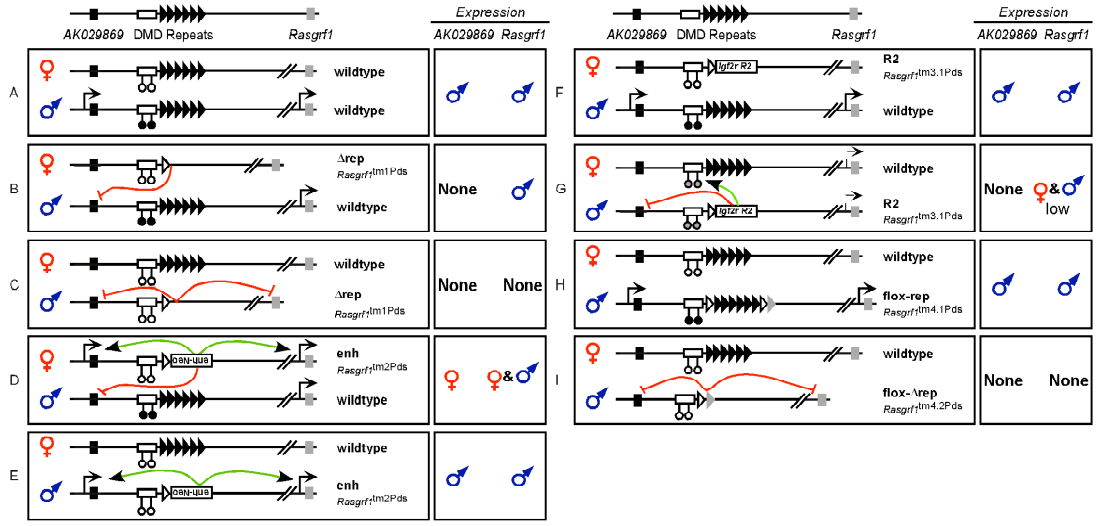


Figure III.4.7. Summary of epigenetic transcriptional control at *Rasgrf1*[8]. Each of the allele combinations represented in the left hand side of panels A-J were tested for their effect on both *AK029869* and *Rasgrf1* expression. Green arrows indicate a positive effect on expression that does not occur in wildtype mice, while red lines ending in bars indicate a negative effect on expression that does not occur in wildtype mice. The ♂ and ♀ symbols in the right hand side of panels A-J indicate whether expression, if any, is from the maternal (♀) or the paternal (♂) alleles.

[8] Figure from Dr. Paul Soloway.

III.4.4.4 Additional ICR mutations result in *trans* silencing

In addition to the repeat-deleted allele, we tested five other ICR mutations for their ability to lead to *trans* silencing of *AK029869* (Figure III.4.5 and Figure III.4.8). The first consists of an extra enhancer (extra enhancer allele) in place of the tandem DNA repeats. When maternally inherited, this mutation results in silencing, in *trans*, of the paternal allele of AK029868 and activation, in *cis,* of the normally silent maternal allele. When paternally inherited, there is no change to the wildtype expression pattern. In either mode of inheritance, there is no change in methylation levels at the DMD. Similarly, an allele that contains the extra enhancer, as well as a loxP-flanked copy of the tandem DNA repeats (extra enhancer/loxP repeats allele), leads to *cis*-activation of the maternal allele and *trans*-silencing of the paternal allele. The third allele behaves exactly as a wildtype allele with the difference that this allele contains a loxP-flanked copy of the tandem DNA repeats (loxP repeats allele). The fourth allele we tested contains both a repeat deletion and an additional Frt site, and behaves as the Δrep allele. The fifth allele contains a DMR referred to as Region 2 from the imprinted and maternally expressed gene, *Igf2r* in place of the *Rasgrf1* tandem DNA repeats (Region 2 allele). When the Region 2 allele is maternally inherited, *AK029869* demonstrates wildtype expression. However, when paternally inherited, the Region 2 allele silences the normally expressed paternal allele in *cis*, which differs from what we observed at *Rasgrf1*. When the Region 2 allele is paternally inherited, *Rasgrf1* expression is biallelic. This is because Region 2 functions as a positive signal for placement of DNA at the *Rasgrf1* DMD, in both *cis* and *trans,* when paternally inherited. One caveat is that paternal *Rasgrf1* exprssion is reduced, but still detectable, with a Region 2 paternal allele. *AK029869*

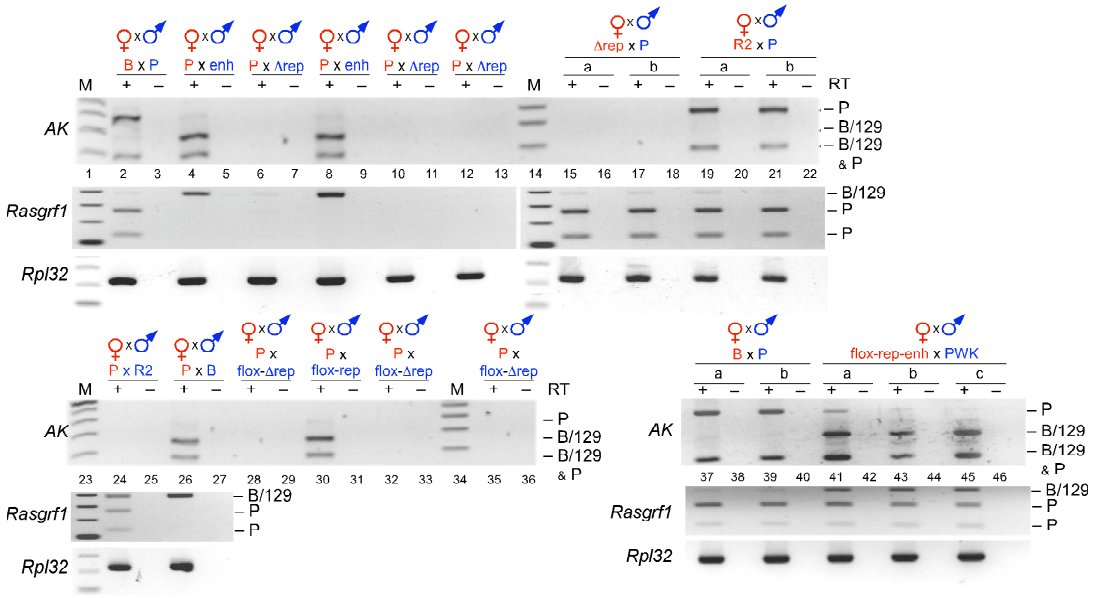


Figure III.4.8. Expression of *AK029869* is affected by several *Rasgrf1* ICR mutations[9]. Gel image showing results of imprinted expression analysis done on crosses between wildtype PWK (P) mice and mice harboring any of a variety of *Rasgrf1* ICR mutations. All of the mutated alleles were on the 129 background, which shares the same AluI restriction polymorphism with the B6 strain. The allele names are consistent with those in Figure III.4.5. Expression of *Rasgrf1* is shown under each gel, as is the positive control amplification for *Rpl32*.

[9] Figure from Dr, Krista Kauppinen.

expression is extremely low, even in wildtype animals, so it is possible that the Region 2 allele reduces *AK029869* expression to below the threshold of detection. Maternal transmission of the Region 2 allele has no effect on DNA methylation at the DMD (Herman 2003).

III.4.4.5 3C analysis

After observing the expression results that occurred with different allele combinations, we wondered whether physical conformation changes could be involved in control of *AK029869* expression. To test this possibility, we did a 3C experiment with a wildtype B6/PWK F1 animal. We used these two strains to take advantage of naturally occurring polymorphisms that allowed us to determine allele-specific interactions after sequencing resulting PCR products. Of the 40 primer pairs tested, we observed evidence of a physical interaction with one of these primer pairs (Figure III.4.9, Figure III.4.10 D). In this case, a PCR product specific for *AK029868* was detected that results when primers located just inside of the BclI sites, and pointing away from each other, are used on cross linked and ligated 3C material (Figure III.4.10 C). Controls lacking either crosslinking or ligation showed no evidence of PCR product. Cloning and sequencing of the PCR product revealed that both intra and inter-allele interactions were occurring (Figure III.4.10 A-C). It is important to keep in mind that these results are very preliminary and need to be repeated and confirmed using additional controls.

III.4.4.6 Effect of timing of repeat deletion on AK029869 expression

To extend this work, I asked when during early development the

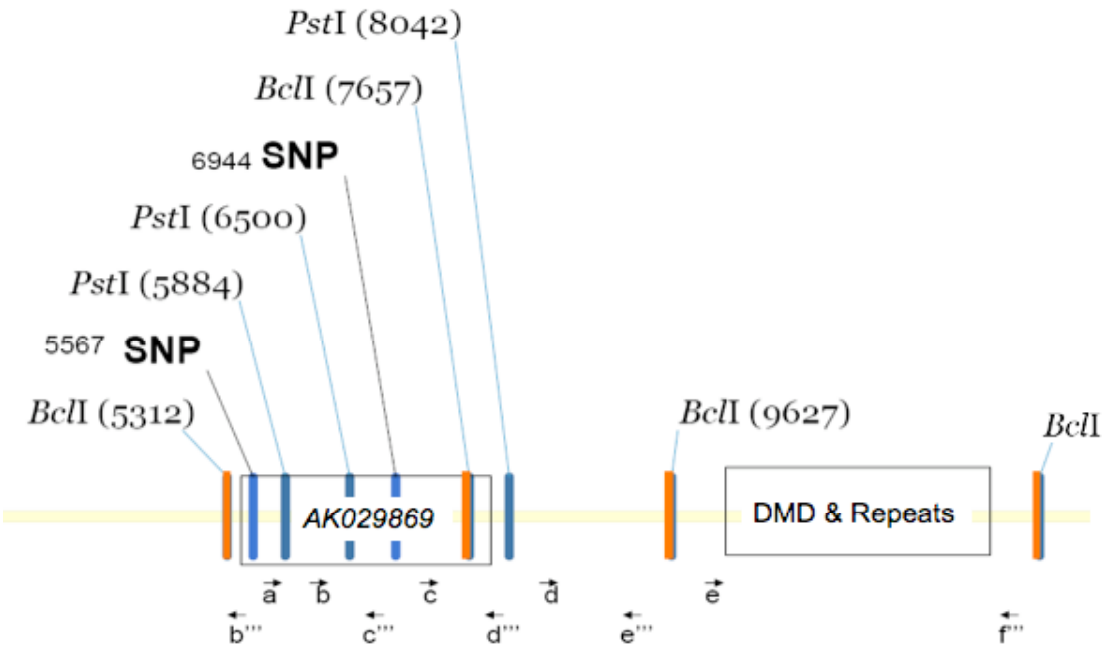


Figure III.4.9. Primers used in the 3C analysis of the *AK029869* locus[10]. Schematic of the *AK029869* transcript and the downstream *Rasgrf1* DMD and DNA repeats. DNA from a B6 x PWK F1 animal was crosslinked and digested with BclI and PstI, then subjected to a low-molarity ligation. Primers labeled a, b, c, d, e, b''', c''', d''', and e''' at the bottom of the figure were used in various combinations on the cross-linked and ligated 3C material. Primers with ' after the primer name indicate that several slightly different primers exist in this location (for example: b''' = b, b', b'1, b'2, and b'3.

[10] Figure from Dr. Krista Kauppinen.

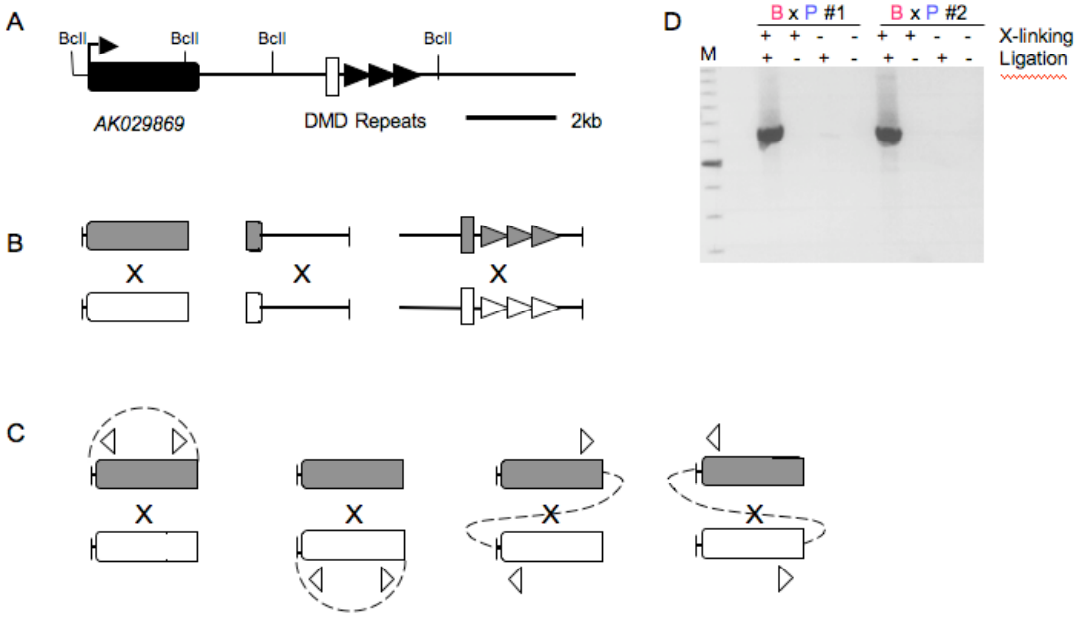


Figure III.4.10 3C analysis indicates that both intra- and inter-chromosomal interactions occur at *AK029869*[11]. (A) A scale schematic showing the location of four BclI sites along the *Rasgrf1* ICR region, including the *AK029869* coding region. The 5' most BclI site is 221bp upstream of the *Ak029869* transcriptional start site. The 3' most site is 447bp downstream of the *Rasgrf1* ICR. (B) Cartoon of the two parental alleles after crosslinking and digestion with BclI. (C) The 5' most BclI fragment is shown after crosslinking, digestion, and ligation. There are four possible ligation conformations, all of which are detectable with nested PCR using the primers (triangles) c (5' – GTG TTT GCA CAC ACG CAT GTG G – 3') and b'2 (5' – GGA GAC ACA GTG ACA TGC AAG GAA G – 3') for 20 cycles followed by c (5' – GTG TTT GCA CAC ACG CAT GTG G – 3') and b' (5' – TAG GGG AGG GCA GCA AAC TG – 3') for 30 cycles. Ligated sites are depicted by dotted grey lines. (D) Gel image showing 3C nested PCR results done on a B6 x PWK F1 animal. Lanes 5-6 and 9-10 are controls lacking cross-linking, while lanes 4, 6, 8, and 10 are controls lacking ligation. Lanes 3 and 7 show the expected PCR product of 749bp only when crosslinked and ligated DNA is used. Sequencing of cloned PCR products revealed that both intra- (8/10) and inter- (2/10) chromosomal interactions occur at *AK029869*.

[11] Figure from Dr. Krista Kauppinen.

repeats are necessary for normal imprinted expression of *AK029869* and to prevent the *trans* effect that occurs with the repeat-deleted allele. To assay allele-specific expression in animals with either the maternal or the paternal allele deleted at specific embryonic time points, I utilized the *Cre*/LoxP system and polymorphic mouse strains. To allow for allele-specific expression analysis, I bred specific *Cre* alleles onto the PWK mouse background. These mice were mated with mice homozygous for a loxP-flanked version of the *Rasgrf1* DNA repeats. In combination with *Cre* recombinase expression, the loxP-flanked repeats can be deleted at specific time points. Zp3 *Cre* is active at e0.0 and deletes the repeats at the one-cell stage, while Meox2 *Cre* is active at e5.5 and deletes the repeats around the time of implantation into the uterine wall. Depending on the direction of the cross, I was able to delete the repeats at these time points on either the maternally or the paternally inherited allele.

From these crosses, I collected P10 brain samples from 50 litters of mice. Within these 50 litters, I genotyped each brain sample for the presence of one 129 allele and one PWK allele at *AK029869* to facilitate allele-specific expression analysis. I also genotyped each brain sample for the degree of deletion of the loxP-flanked copy of the *Rasgrf1* repeats. Within my samples, I had brains from +/e0.0ΔRepeats animals, e0.0ΔRepeats/+ animals, +/e5.5ΔRepeats animals, and e5.5ΔRepeats/+ animals. I prepared cDNA from all samples showing deletion of the repeats and carrying one 129 and one PWK allele at AK029869. Then, I performed expression PCR on the cDNA samples and digested the resulting PCR products with AluI. AluI yields different restriction banding patterns due to polymorphisms between the 129

and PWK mouse strains, allowing for allele specific expression analysis.

In each case, there are a couple of samples that deviate from the patterns described. However, a general pattern does emerge from the data. When the DNA repeats are deleted at e0.0 with *Zp3 Cre* and inherited paternally, expression of *AK029869* switched from paternal expression to maternal expression (Figure III.4.11 A). No data are available for DNA repeats deleted at e0.0 and inherited maternally. I set up a breeding box with *Prm Cre* males crossed to females homozygous for loxp-flanked copies of the *Rasgrf1* repeats, but no pups were born before I switched projects. When the DNA repeats are deleted at e5.5 with *Meox2 Cre* and inherited paternally, expression of *AK029869* switched from paternal expression to null expression (Figure III.4.11 B). Similarly, when the DNA repeats are deleted at e5.5 with *Meox2 Cre* and inherited maternally, expression of *AK029869* switched from paternal expression to null expression (Figure III.4.11 C).

III.4.4.7 Preliminary FISH assay for physical proximity of the two AK029869 homologues

I also began to explore the possibility that the DNA repeats are needed to mediate physical interaction between the two parental copies of chromosome 9 by fluorescent *in situ* hybridization (FISH). Preliminary FISH studies were conducted on wildtype MEFs and MEFs homozygous for a deletion of the *Rasgrf1* tandem DNA repeats. In this experiment, the *Rasgrf1* BAC RP24-228H14 was labeled with or Spectrum orange and was hybridized to either MEFs of both genotypes. It is important to note that the results obtained from this experiment are very preliminary. Nevertheless, the two

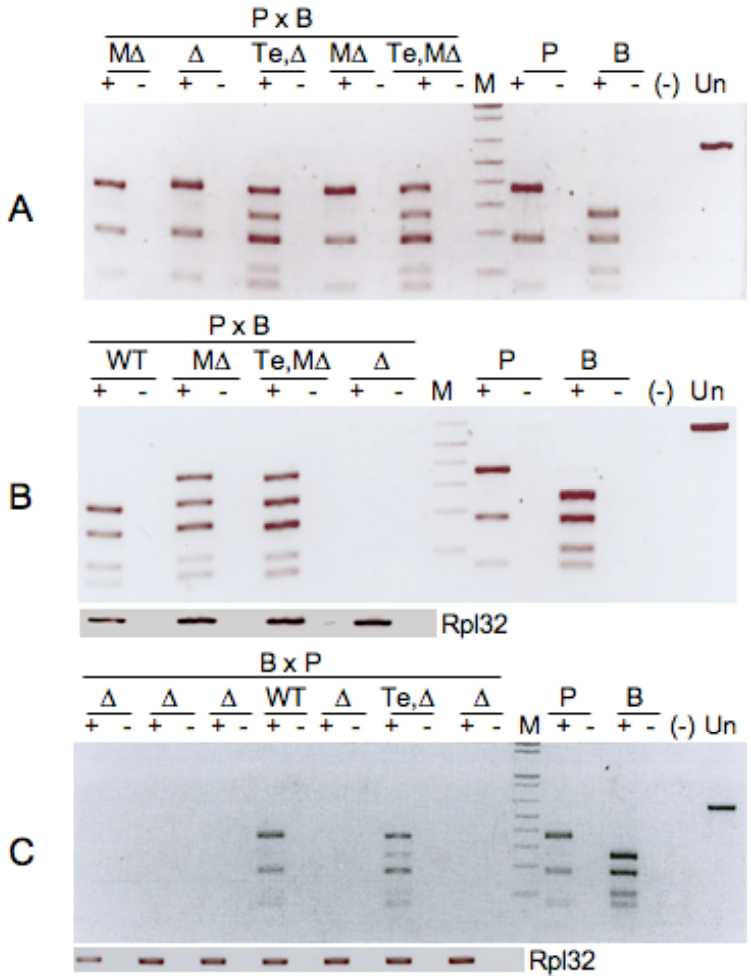


Figure III.4.11. *AK029869* expression analysis of progeny carrying deletions of the *Rasgrf1* tandem DNA repeats. Brains or testis (Te) of P10 progeny carrying either maternally (C) or paternally (A, B) inherited deletions of the *Rasgrf1* tandem DNA repeats were analyzed for the expression status of *AK029869*. The tandem repeats were deleted at either e0.0 (A) with *Zp3 Cre* or at e5.5 (B, C) with *Meox2 Cre*. PxB = PWK mother rossed to a PWK father. BxP = B6 mother crossed to a PWK father. Δ = complete deletion of the tandem DNA repeats. MΔ = mosaic deletion of the tandem DNA repeats. WT = wildtype. P = PWK control. B = B6 control. (-) = water. Un = uncut. + = reverse transcriptase added. - = no reverse transcriptase added. In cases of null expression, an *Rpl32* control was done to ensure RNA and cDNA were properly prepared (B and C, *Rpl32*).

chromosome 9 homologues identified by the DNA FISH probe appear to be closer together in the wildtype MEFs as compared to the MEFs homozygous for a deletion of the *Rasgrf1* DNA repeats (Figure III.4.12). These results appear to support the initial results from the 3C analysis (Figure III.4.10). Although, as noted above, additional work needs to be done to confirm the 3C results. Moreover, additional FISH analysis needs to be done with a greater sample size andstatistical analysis to confirm this observation.

III.4.5 Discussion

This trans-expression effect constitutes the second of three observed *trans* expression effects within the *Rasgrf1* imprinted cluster on mouse chromosome 9. Although additional samples need to be analyzed to confirm the results of the deletion analysis, we know that the *Rasgrf1* DNA repeats are important for proper imprinted expression of *AK029869.* It appears that deletion of the DNA repeats before implantation leads to a reversal of imprinted expression of *AK029869*. At *Rasgrf1*, deletion of the repeats only affects imprinted expression until the point of implantation, although the pattern of perturbed expression is consistent. However, at *AK029869*, it seems that deletion of the repeats after e5.5 does have an effect on expression. Both a maternally and a paternally inherited deletion of the repeats leads to complete silencing of *AK029869* when deleted at e5.5. Of course, more samples are needed to confirm the initial results, and no data are currently available concerning the effect of a paternally inherited e0.0 deletion.

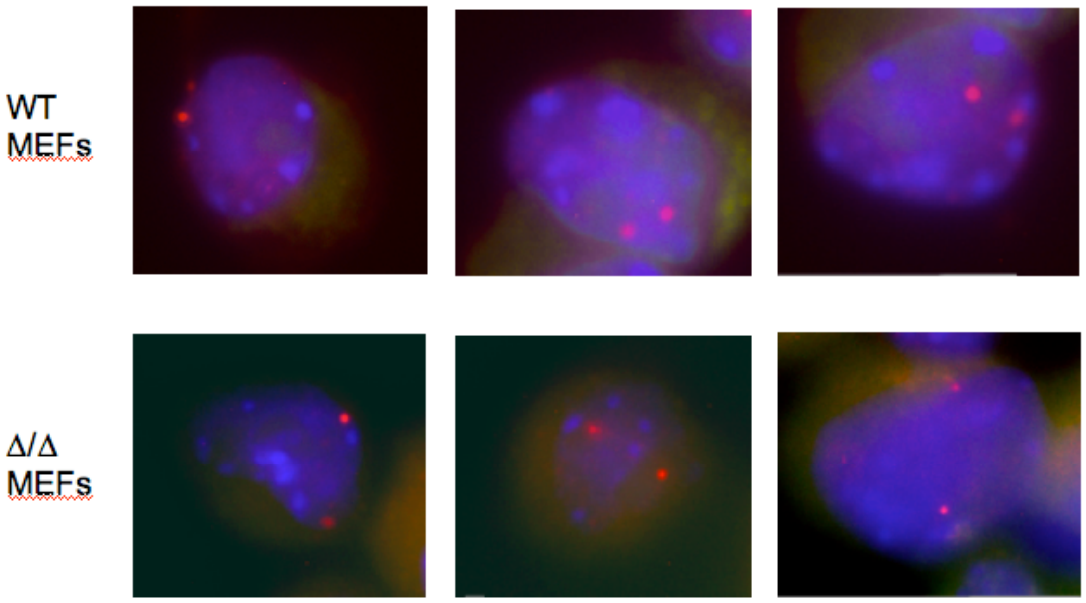


Figure III.4.12. FISH analysis of physical proximity of the two *AK029869* alleles. FISH using *Rasgrf1* BAC probe RP24-228H14 labeled with Spectrum orange. Hybridization was done on both wildtype (WT) MEFs and MEFs homozygous for a deletion of the *Rasgrf1* tandem DNA repeats. Images shown are a merge of the red channel, which captures the *Rasgrf1* BAC probe, and the DAPI channel, which captures DNA. Red spots in each image represent the two homologues of mouse chromosome 9. Each image depicts one individual cell at 100X magnification under oil.

Regarding how the expression of *AK019869* is regulated, we considered several possibilities. We first considered that *AK029869* might be regulated by DNA methylation levels at the DMD. However, maternal inheritance of a repeat-deleted allele, an extra enhancer allele, or an extra enhancer/loxP repeats allele all preserve DNA methylation at the paternal DMD, but lead to silencing in *trans* of the paternal allele (Figure III.4.7). Furthermore, paternal inheritance of an extra enhancer allele preserves DNA methylation at the paternal DMD, but leads to silencing of the paternal allele in *cis.*

We also considered that imprinted expression of *AK029869* may require both the maternal and the paternal copies of the tandem DNA repeats. To the contrary, mice inheriting a paternal copy of an extra enhancer/loxP repeats allele have both the maternal and the paternal tandem DNA repeats, but lack expression of *AK029869,* while mice with a maternally inherited Region 2 allele lack one copy of the DNA repeats but express *AK029869* from the paternal allele (Figure III.4.7).

In addition, we considered that sequence spacing within the ICR might be critical for proper expression of *AK029869*. Except, paternal inheritance of the Region 2 allele retains wildtype sequence spacing, as Region 2 and the tandem DNA repeats are both approximately 2kb, but leads to silencing in *cis* of the paternal allele. In contrast to the binary switch model for expression of *Rasgrf1* (Figure III.4.2), there seemed to be no clear mechanism of regulation at *AK029869.*

However, there are three patterns that emerge from these expression data. First, any changes in spacing, relative to wildtype, on the maternal allele lead to silencing in *trans* of the paternal allele (Figure III.4.7). For example, the Region 2 allele contains an approximately 2kb deletion of the tandem DNA repeats but, since Region 2 is roughly 2kb, insertion of Region 2 retains normal sequence spacing. When maternally inherited, the Region 2 allele allows expression of the paternal allele. On the other hand, alleles containing an extra enhancer/loxP repeats change the sequence spacing of the region, and lead to paternal allele silencing. The one exception is the extra enhancer allele, which keeps the maternal allele spacing but silences the paternal allele in *trans* (Figure III.4.7).

Second, deletion of the paternal repeats, regardless of the resulting sequence spacing, leads to silencing in *cis* of the paternal allele (Figure III.4.7). Again, the one exception occurs if the repeats are replaced with the extra enhancer (Figure III.4.7). For example, paternal inheritance of the Region 2 allele deletes the tandem DNA repeats and preserves sequence spacing, but silences the paternal allele. The extra enhancer allele deletes the tandem DNA repeats but preserves sequence spacing and allows expression of the paternal allele.

Finally, any allele with an extra enhancer leads to activation of that allele in *cis*. Replacement of either the paternal or the maternal tandem DNA repeats with the extra enhancer leads to *cis* activation of that allele, regardless of sequence spacing or the presence of the repeats on that allele. This explains the exception to the first and second patterns discussed above.

Thus, it appears that sequence spacing in the region of the repeats, but not presence of the repeats, on the maternal allele is important, while the presence of the repeats, but not sequence spacing, on the paternal allele is important. Also, the presence of an extra enhancer may override these two points to allow expression in *cis* but not in *trans*, indicating that access to an enhancer is necessary for expression of *AK029869*.

These observations support the possibility of conformational regulation, indicated by our preliminary 3C results. The 3C experiment performed by Dr. Krista Kauppinen indicates that there may be an inter-chromosomal interaction occurring at *AK029869* in wildtype animals (Figure III.4.9). Additionally, preliminary FISH studies may support the results from the 3C analysis (Figure III.4.10). In this experiment, the *Rasgrf1* BAC RP24-228H14 was labeled with or Spectrum orange and was hybridized to either wildtype MEFs or to MEFs homozygous for a deletion of the *Rasgrf1* tandem DNA repeats. The two chromosomes identified by the DNA FISH probe appear to be closer together in the wildtype MEFs as compared to the MEFs homozygous for a deletion of the *Rasgrf1* DNA repeats. However, additional analysis needs to be done with a greater sample size and statistical analysis to confirm this observation.

If conformational regulation is occurring, it could be dependent on sequence spacing, repeat-content, or a combination of the two, which would allow or excludes access to an enhancer. Conformational regulation of imprinted expression has been shown to operate at the silent maternal allele of *Igf2* (Qiu 2008; Ling 2006; Kato 2005; Murrell 2004). The silent maternal *Igf2* allele forms a complex three-dimensional loop that prevents enhancers

from interacting with the Igf2 promoters. This allows enhancers to have direct access to the Igf2 promoters on the expressed paternal allele of *Igf2*. The complex loop on the maternal allele is formed by interactions involving a DMD, the ICR, and enhancers, all of which are implicated in imprinted control at *Rasgrf1.* Additionally, binding of CTCF to the unmethylated maternal ICR (which also occurs at *Rasgrf1)* cooperates with interchromosomal interactions to create a barrier, which blocks the access of all enhancers to *Igf2*, thereby silencing the maternal *Igf2.* Our results indicate that imprinted expression of *AK029869* may be regulated by similar conformational mechanisms. The results discussed above may be indicative of the first evidence of physical inter- or intra-chromosomal interactions within the *Rasgrf1* imprinted domain.

III.5 DEVELOPMENTAL TIMING OF THE PLACEMENT OF DNA METHYLATION AND H3K27me3

III.5.1 Abstract

At the imprinted *Rasgrf1* locus, the histone modification H3K27me3 are mutually exclusive epigenetic marks (Lindroth 2008). As this is a very recent finding, relatively little is known about the timing of the placement of H3K27me3, relative to the timing of the placement of DNA methylation. There are two aims for this project, both of which seek to identify the timing with which aberrant epigenetic marks are placed at *Rasgrf1*. The first goal is to determine whether the appearance of H3K27me3 coincides with a loss of DNA methylation at the endogenous *Rasgrf1* locus. At the endogenous *Rasgrf1* locus, DNA methylation is controlled by a series of tandem DNA repeats, which provide a positive signal to methylate the paternally inherited DMD (Yoon 2002). Using a system in which we can temporally control the loss of DNA methylation, we will perform ChIP to determine whether the appearance of H3K27me3 coincides with a loss of DNA methylation. The second goal uses a transgenic system in which aberrant DNA methylation appears on a *Rasgrf1* transgene. This will allow us to identify the time at which the aberrant methylation appears.

III.5.2 Introduction

Rasgrf1 is a paternally expressed imprinted gene on mouse chromosome 9. Expression of *Rasgrf1* is controlled by a binary switch comprised of a differentially methylated domain (DMD) and a series of DNA

repeats, located 30kb upstream of the *Rasgrf1* promoter. The DNA repeats provide a positive signal for placement of DNA methylation at the DMD, which leads to expression of *Rasgrf1* from the paternal allele in neonatal brain. Using an allele with loxP-flanked DNA repeats and various Cre recombinases, which remove the repeats at different times during development, the DNA repeats were shown to be critical for this methylation to occur in the period between e0.0 and e5.5 (Holmes 2006). However, after e5.5 (when genomic methylation reprogramming has finished) the DNA repeats are dispensible.

Additionally, a transgene containing the *Rasgrf1* DMD and DNA repeats behaves differently based on its mode of inheritance (Figure III.5.1 A; Herman and Park, unpublished). A purely maternal lineage demonstrates an unmethylated DMD, as expected, and a purely paternal lineage demonstrates a methylated DMD, also as expected. However, passage of the transgene from the paternal germline and then back through the maternal germline leads to unexpected results in the progeny of this cross (Figure III.5.1 B). Instead of proper reprogramming to the unmethylated state, which is expected of a maternally inherited transgene, the progeny demonstrate an improperly methylated maternal copy of the transgene. We know that germline reprogramming occurs properly, as ooctyes contain an unmethylaed copy of the transgene, but that somatic maintenance of the transgenic DMD in an unmethylated state fails (Figure III.5.2; Park, unpublished). Furthermore, we know that this occurs sometime between e0.0 and e7.5. To date, we have no information regarding the status of histone modifications at the transgenic DMD.

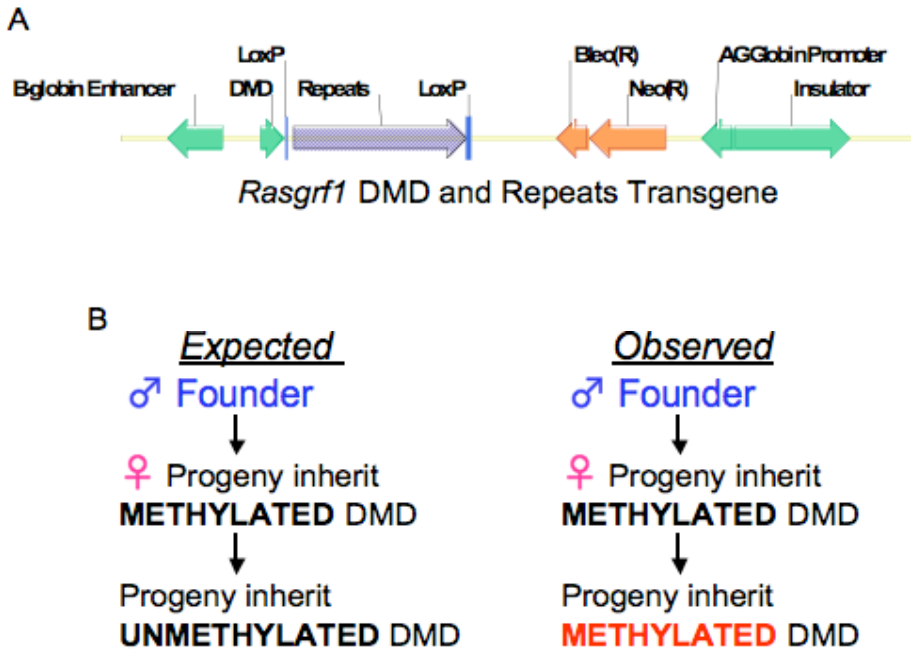


Figure III.5.1. A minimal *Rasgrf1* transgene behaves differently based on its mode of inheritance[12]. (A) Cartoon of a minimal *Rasgrf1* transgene containing the *Rasgrf1* DMD and loxP-flanked tandem DNA repeats. The DMD and repeats are located between the human A-gamma globin promoter and the beta-globin enhancer. Also contained on the transgene is a Neo cassette, between the repeats and the A-gamma globin promoter. Downstream of the A-gamma globin promoter is an insulator sequence. (B) The *Rasgrf1* minimal transgene behaves differently based on its mode of inheritance (Herman and Park, unpublished). The left hand pedigree showing purely maternal inheritance demonstrates an unmethylated DMD, as expected. The pedigree on the right depicts passage of the transgene from the paternal germline and then back through the maternal germline. This mode of inheritance leads to unexpected results in the progeny of this cross. Instead of proper reprogramming to the unmethylated state, which is expected of a maternally inherited transgene, the progeny demonstrate an improperly methylated maternal copy of the transgene.

[12] Methylation analysis by Dr. Herry Herman.

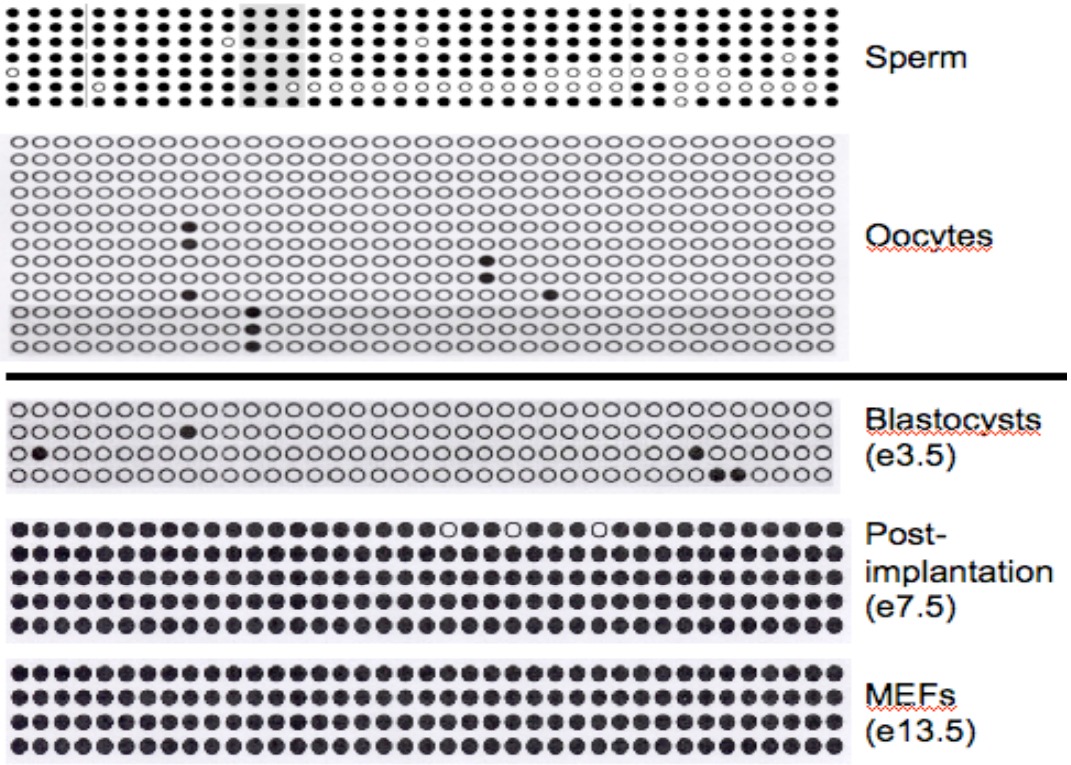


Figure III.5.2. The minimal *Rasgrf1* transgene is properly reprogrammed upon passage through the maternal germline[13]. Bisulfite sequencing of oocytes demonstrates proper reprogramming to the unmethylated state does occur at the transgenic DMD (Park, unpublished). Bisulfite sequencing of later stages of the fertilized embryo revesla that somatic maintenance of the transgenic DMD in an unmethylated state fails sometime between e0.0 and e7.5 (Park, unpublished). Each circle represents an individual CpG dinucleotide along the transgenic DMD. Black circles represent methylated CpGs and white circles represent unmethylated CpGs.

[13] Bisulfite sequencing by Dr. Yoon Jung Park.

At the endogenous *Rasgrf1* locus, DNA methylation and the histone modification H3K27me3 are antagonizing epigenetic marks (Figure III.5.3; Lindroth 2008). I would like to determine whether the appearance of H3K27me3 coincides with a loss of DNA methylation at the endogenous *Rasgrf1* locus, and I would also like to narrow down the time point at which the inappropriate placement of DNA methylation occurs on the *Rasgrf1 transgene*. Using the same loxP-flanked repeat allele and various *Cre* recombinase alleles, I planned to use embryonic fibroblast cells and ChIP to determine whether the appearance of H3K27me3 coincides with a loss of DNA methylation. If so, I expect that when the DNA repeats are deleted at e0.0, H3K27me3 will be present on the unmethylated paternal allele and when the DNA repeats are deleted at e5.5, H3K27me3 will be absent on the methylated paternal allele.

I used mice carrying a *Rasgrf1* transgene that has been passed through various modes of inheritance to narrow down the time point at which the inappropriate placement of DNA methylation occurs. These modes of inheritance include: pNIDR4 line 2771 only paternally inherited (appropriately methylated in somatic tissue), pNIDR4 line 3047 only maternally inherited (appropriately unmethylated in somatic tissue), pNIDR4 line 2771 paternal inheritance followed by maternal inheritance (appropriately unmethylated in somatic tissue), pNIDR4 line 3047 only maternal inheritance followed by paternal inheritance (appropriately methylated in somatic tissue), and pNIDR4 line 3047 only maternal inheritance followed by paternal inheritance followed again by maternal inheritance (**inappropriately** methylated in somatic tissue) (Table III.5.1). Progeny from these crosses will be analyzed for DNA

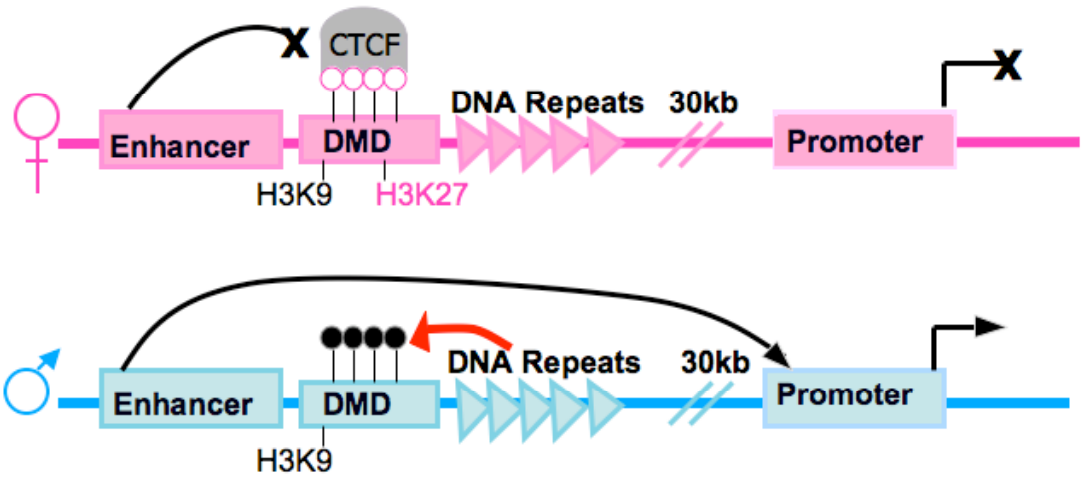


Figure III.5.3. Current model for epigenetic control of imprinted expression at *Rasgrf1.* One the paternal allele (depicted in blue), a series of DNA repeats are responsible for the establishment and maintenance of paternal allele-specific DNA methylation at CpG dinucleotides in the DMD (Yoon 2002; Holmes 2006). Placement of DNA methylation at the DMD prevents CTCF binding and allowing promoter-enhancer interaction, resulting in transcriptional activation (Yoon 2005). Placement of DNA methylation at the DMD also blocks the binding of H3K27me3 (Lindroth 2008). The maternal allele (depicted in pink) is unmethylated at CpG dinucleotides in the DMD. The unmethylated status if the DMD allows binding of CTCF and prevents the *Rasgrf1* upstream enhancer from interacting with the promoter, resulting in suppression of transcription (Yoon 2005). Lack of methylation at the DMD also allows placement of the silencing mark H3K27me3 (Lindroth 2008). H3K9me3 is found on both the maternal and the paternal allele and therefore does not appear to control differential expression. Circles above the DMD represent the methylation status of CpG dinucleotides within the DMD. Black circles represent a methylated DMD and white circles represent an unmethylated DMD.

methylation at e2.5, e3.5, and e7.5. I would also like to use an additional mating involved pNIDR4 line 3047 maternal inheritance only/*Suz12*+/- females crossed to *Suz12*+/- males. This cross will yield 3047 maternal inheritance/*Suz12*-/- embryos, which will lack H3K27me3 and may inappropriately gain DNA methylation. Progeny from this cross will be analyzed for DNA methylation at e7.5 only, as each embryo needs to be genotyped individually to ensure inheritance of both the pNIDR4 transgene and two *Suz12* knockout alleles.

III.5.3 Materials and methods

III.5.3.1 Materials collection

MEFs were created from timed between female mice carrying either *Zp3 Cre* or *Meox2 Cre* alleles and male mice homozygous for loxP-flanked copies of the *Rasgrf1* DNA repeats. Females were placed into breeding cages with males overnight and evidence of plugging was checked each morning. At e13.5, female mice were sacrificed and embryos were dissected out in a laminar flow hood. The head and internal organs were removed and organ samples were saved for genotyping. Each embryo was minced in 0.25% trypsin-EDTA and incubated for 20 minutes at 37C in a 5% CO2 incubator. Trypsinization was quenched with EF media (1.2g/L $NaHCO_3$, 6.24g/L HEPES, 75mls heat inactivated fetal calf serus, 5mls 100X Gibco no. 11140 non-essential amino acids, 5mls 100X Gibco no. 15070 antibiotic mix, and 0.4mls of a 1:100 dilution of beta–mercaptoethanol in 450mls Gibco no. 12800 DMEM), the suspension was moved to a 50ml conical tube, and large particles

Table III.5.1. Summary of all materials collected for minimal *Rasgrf1* transgene methylation timecourse. The top two genotypes are heterozygous for a loxP-flanked copy of the *Rasgrf1* DNA repeats at the endogenous site in the genome. These are controls for methylation analysis, as their methylation state is known and primers designed to the minimal transgene will amplify the loxP-flanked allele. The remaining six genotypes are experimental genotypes.

Genotype of embryos	Expected adult somatic Methylation	Number of e2.5 embryos collected	Number of e3.5 embryos collected	Number of e7.5 embryos collected
+/pYJC6 (control)	Paternal	8	11	9
pYJC6/+ (control)	Paternal	9	12	8
2771 Pat	Methylated	18+22+15	17+9+10	5
2771 Pat, Mat	Unmethylated	9	8+9	8
3047 Mat	Unmethylated	10+15	12	12+8
3047 Mat, Pat	Methylated	12	10+8	9
3047 Mat, Pat, Mat	Methylated	14	11+7	11
3047 Mat, Suz12-/-	Methylated	n/a	n/a	5

were allowed to settle out for 5 minutes. The suspension was removed from the settled material to a new tube, spun at 1.5g for 5 minutes, and plated to a T75. Each embryo was genotyped to confirm deletion of the repeats.

Early embryos were collected from timed reciprocal matings between mice homozygous for loxP-flanked copies of the *Rasgrf1* DNA repeats and wildtype mice, as well as from timed matings between mice carrying a *Rasgrf1* transgene (pNIDR4) and wildtype mice. The latter matings were done in various inheritance patterns: pNIDR4 line 2771 only paternally inherited, pNIDR4 line 3047 only maternally inherited, pNIDR4 line 2771 paternally inheritance followed by maternal inheritance, pNIDR4 line 3047 only maternal inheritance followed by paternal inheritance, and pNIDR4 line 3047 only maternal inheritance followed by paternal inheritance followed again by maternal inheritance. The final mating involved pNIDR4 line 3047 maternal inheritance only/*Suz12*+/- females crossed to *Suz12*+/- males. This cross will yield 3047 maternal inheritance/*Suz12*-/- embryos. Females were placed into breeding cages with males overnight and evidence of plugging was checked each morning. At three developmental stages, e2.5, e3.5, and e7.5, embryos were dissected out and DNA prepared for methylation analysis. For the pNIDR4 line 3047 maternal inheritance/*Suz12*-/- crosses, embryos were only collected at e7.5. After DNA preparation, these embryos were genotyped for the presence of pNIDR4 and two *Suz12* deletion alleles (Figure III.5.4). In each case, embryos were pooled from multiple timed matings to yield enough material for analysis.

III.5.3.2 DNA isolation

For MEFs, during embryo dissection, a tail snip and hindlimb were reserved and lysed with filter-sterilized Laird's lysis buffer (50 mM pH 8.0 Tris, 100 mM EDTA, 100mM NaCl, 1% SDS, and 100ug/mL proteinase K) overnight at 55C with shaking at 220rpm. DNA was then prepared by isoproanol precipitation for genotyping. For the early embryo samples, genotyping was not done because of the way the crosses were set up, except for the *Suz12* animals. In this case, DNA was prepared by lysing individual embryos in filter-sterilized Laird's buffer (50 mM pH 8.0 Tris, 100 mM EDTA, 100mM NaCl, 1% SDS, and 100ug/mL proteinase K) overnight at 55C with shaking at 220rpm. Then, DNA was precipitated with 3ul of glycoblue, 1/10 volume NaOAc, pH4, and 1 volume of isporopanol. Embryo DNAs of the correct genotype were pooled after genotyping.

III.5.3.3 Genotyping

The organ samples from the embryos used to make MEFs were genotyped using PDS16 (5'- GCA CTT CGC TAC CGT TTC GC -3'), PDS17 (5'- TGT CCT CCA CCC CTC CAC C -3'), PDS18 (5'- TTT CTG CCA TCA TCC CAG CC -3') to detect the presence of a repeat-deleted paternal allele. Cycling conditions were 94C for 10sec, 61C for 20 sec, 72C for 50 sec for 40 cycles. MEFs from those embryos with a paternally inherited repeat deletion were expanded for ChIP.

DNA from the pNIDR4 line 3047 maternal inheritance/*Suz12*-/- crosses were genotyped for the presence of pNIDR4 and two *Suz12* deletion alleles (Figure III.5.4). To detect pNIDR4, primers PDS16 (5'- GCA CTT CGC TAC

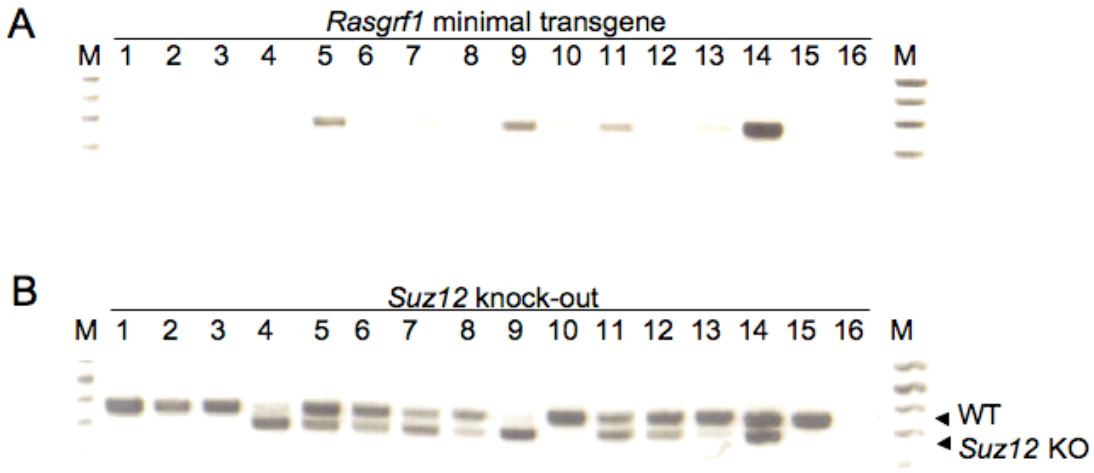


Figure III.5.4. Genotyping of individual e7.5 embryos for the presence of both the *Rasgrf1* minimial transgene and double knockout of *Suz12.* (A) PCR done on individual e7.5 embryos using primers specific to the *Rasgrf1* minimal transgene. Embryo DNAs in lanes 5, 9, and 11 are transgene positive. Lane 14 is a positive control, lane 15 is a B6 genomic DNA negative control, and lane 16 is water. (B) (A) PCR done on individual e7.5 embryos using primers specific to *Suz12*. *Suz12* wildtype alleles result in a 350bp band. *Suz12* knockout alleles result in a smaller 250bp band. Embryo DNAs in lanes 4 and 9 are *Suz12-/-*. Lane 14 is a heterozygous knockout positive control, lane 15 is a B6 genomic DNA control, and lane 16 is water.

CGT TTC GC -3') and pYJC6R2 (5'- CCT GCA GGT CGA CAT AAC TTC -3') were used under the conditions 94C for 10sec, 61C for 20sec, 72C for 50sec for 40 cycles. To detect *Suz12* deletion, primers PDS674 (5'- TGG AGC TGG AGT TAC CTG -3'), PDS675 (5' GCC TGA AGA ACG AGA TCA -3'), and PDS676 (5'- CCA GGT CAT CTT GTG GAG -3') were used under the conditions 94C for 10sec, 57C for 45sec, 72C for 30sec for 36 cycles.

V6.4 Results

At this point, I have created MEF lines with the repeats deleted at both e0.0 and at e5.5 (Figure III.5.5). These cell lines are ready for expansion and ChIP analysis for the presence of H3K27me3. I have also collected embryos at three developmental stages (e2.5, e3.5, and e7.5) a variety of mouse lines for methylation analysis (Table III.5.1). I have embryos from two control lines: loxP-Repeat x +/+ and +/+ x loxP-Repeat as well as a variety of experimental lines. Experimental lines include: Paternal only Tg x +/+, Pat. followed by Maternal Tg x +/+, Maternal only Tg x +/+, Mat. followed by Paternal Tg x +/+, Mat. followed by Pat. followed by Mat Tg x +/+, and Maternal only Tg/*Suz12-/-*. I have prepared embryo genomic DNA from all of these lines and have designed several sets of transgene (and loxP-Repeat) specific primers for methylation analysis using sequenom MASSarray.

V6.5 Discussion

I have collected a sufficient number of embryos from each of the eight genotypes, at three developmental stages for DNA methylation analysis (Table

III.5.1). DNA methylation analysis can be done on these DNAs by bisulfite treating as previously described (Yoon 2002 and in section III.1.6.3). Primers PDS638 (5' – TAG TAG TAG TGG TTG GGG TAG GGG TAG T – 3') and PDS641 (5' – ACA AAA TAC CAA TAA AAA TCT ACA ATA AAT TC – 3') can be used to PCR amplify the bisulfite treated DNA. PCR products can then be cloned using Topo-TA cloning and individual colonies can be sequenced for methylation analysis. Alternatively, primers PDS638 and PDS641 can be modified to contain a T7 promoter sequence at the 5' end of one primer for Sequenom MASSarray DNA methylation analysis of bisulfite treated DNAs. These DNAs are stored in the enzyme -20C freezer and are in a box labeled "pNIDR4 Reprogramming".

I have also created MEF lines that have the *Rasgrf1* DNA repeats deleted at the endogenous locus at e0.0 (*Zp3 Cre* lines*)* and e5.5 (*Meox2 Cre* lines). The MEF lines are entered into the LN2 database and are stored in the liquid nitrogen tank. I also have cell samples at $1x10^7$ that have been crosslinked but not sonicated and cell samples that have been both crosslinked and sonicated store d in the -80C freezer in a box labele "ChIP". These samples are ready for ChIP analysis of H3K27me3 binding using Millipore/Upstate item 07-449 (lot 24440 has worked well in the past). This can be done using primers PDS16 (5' – GCA CTT CGC TAC CGT TTC GC - 3') and PDS24 (5' – CTT GCA GTT CGA CAT AAC TTC - 3') under the conditions 94C for 30 seconds, 62C for 30 seconds, and 72C for 50 seconds for 40 cycles.

III.6 EFFECTS OF A *DCR* KNOCK DOWN ON *TRANS*-ALLELE METHYLATION AND EXPRESSION

This project was done in cooperation with Dr. Yoon Jung Park and Dr. Anders Lindroth. Dr. Park began the multi-generation crosses necessary to obtain five alleles in a single animal. She also characterized the RC allele *trans*-effect and provided Figure III.6.2. Dr. Lindroth created the *Dcr* conditional knockdown animals and provided Figure III.6.1, Figure III.6.3, and Figure III.6.4.

III.6.1 Abstract

Conditional knockdown of the small RNA processing enzyme, *Dcr,* leads to a reduction in the levels of DNA methyltransferases (Figure III.6.1; Benetti 2008; Lindroth, unpublished). Additionally, when a BAC transgene allele (RC allele) containing the entire 250kb *Rasgrf1* genomic region is used in combination with a *Rasgrf1* allele where the repeats have been replaced with the ICR from a maternally expressed imprinted gene, *Igf2r*, the normally silent and unmethylated RC allele is both methylated and expressed when inherited maternally (Figure III.6.2; Park, unpublished). This *trans* methylation, and corresponding expression, effect can be abolished with a knock down in the levels of *Dcr*. However, the effect is variable and may be dependent upon the level of *Dcr* knockdown. This suggests that once *Dcr* expression is reduced past a certain threshold, *trans* methylation of RC either does not occur, or occurs at levels so low that RC is not expressed.

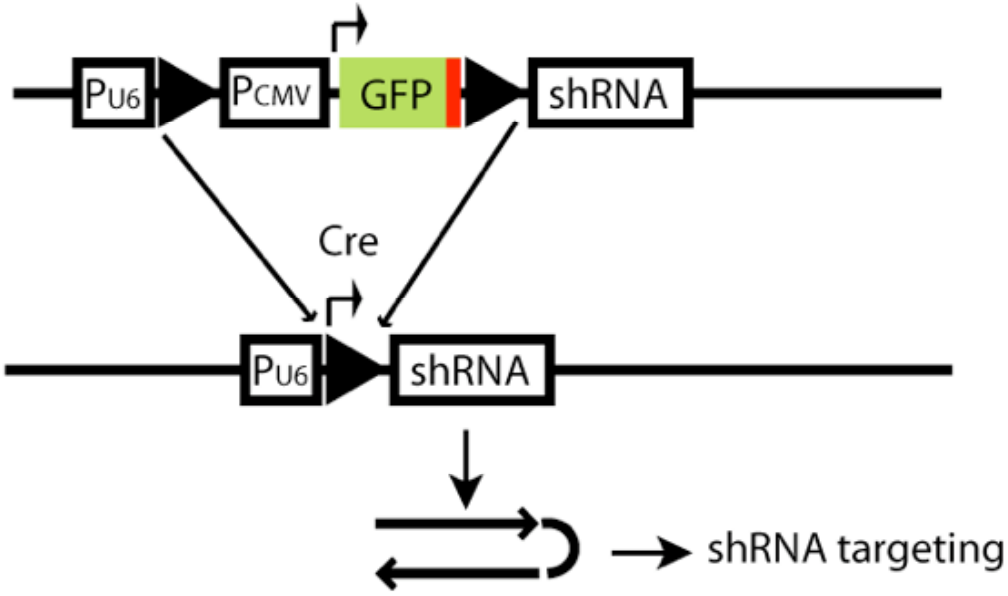
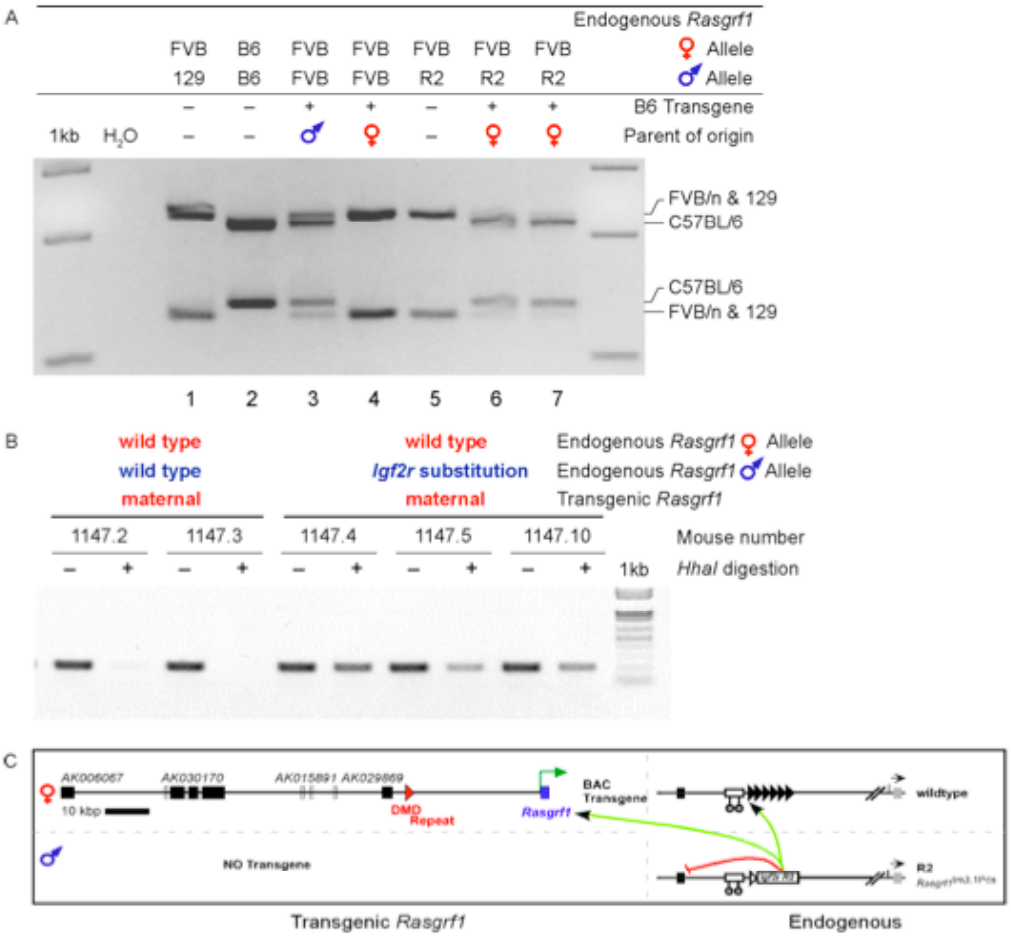


Figure III.6.1. Cartoon depicting the conditional knockdown allele of the small RNA processing enzyme, *Dcr*[14]. The *Dcr* conditional knockdown allele contains a GFP reporter under the control of the *CMV* promoter for screening. The GFP cassette is flanked by loxP sites for removal with *Cre* recombinase. Once deleted, an upstream Pu6 promoter is able to drive expression of a short hairpin RNA with homology to *Dcr*, resulting in knockdown of *Dcr* expression levels.

[14] Figure from Dr. Anders Lindroth.

Figure III.6.2. Maternally transmitted RC transgene exhibits a *trans*-effect identical to that at the endogenous locus[15]. All crosses used mice from the FVB/n background with a C57BL/6 (B6) RC transgene and mates from different strain backgrounds that were either wildtype at the *Rasgrf1* locus or carried a paternally inherited Region 2 allele. (A) RT-PCR analysis of brain RNA from control crosses with wildtype mates confirmed that the RC1 transgene is expressed only upon paternal transmission (lanes 1- 4). RC females were crossed to males carrying the Region 2 allele, which replaces the paternal methylation promoter at *Rasgrf1* with the Region 2 imprinting control region from *Igf2r*. The paternally transmitted Region 2 mutation at the endogenous locus activates the unlinked and normally silent maternal RC transgene in the progeny (lanes 6-7). This effect is identical to the *trans*-effect at the endogenous locus (Herman 2003). Note that Region 2 is from strain 129 and that 129 and FVB/n share a polymorphism that differs from B6. (B) DNA methylation analysis of the aberrantly activated maternal RC transgene was performed using DNA that was undigested (-) or previously digested with the methylation sensitive enzyme HhaI prior to amplification. Primers used were specific for the RC transgenic DMD and spanned four HhaI sites in the DMD. In this assay, the presence of PCR product in the HhaI digested lanes indicates the presence of DNA methylation over the region being analyzed. PCR product is visible in the digested DNA lanes only when the transgene is combined with a paternally inherited Region 2 allele. This is indicative of DNA methylation on the maternal RC. (C) A model depicting *trans*-effects controlled by the paternal Region 2 allele, as reported earlier at the endogenous locus and at the unlinked RC transgenic locus (Herman 2003).

[15] Figure and caption adapted from Yoon Jung Park's thesis, 2008.


A
Endogenous Rasgrf1
FVB B6 FVB FVB FVB FVB FVB
129 B6 FVB FVB R2 R2 R2
♀ Allele
♂ Allele
– – + + – + + B6 Transgene
1kb H2O – – ♂ ♀ – ♀ ♀ Parent of origin
FVB/n & 129
C57BL/6
C57BL/6
FVB/n & 129
1 2 3 4 5 6 7
B
wild type
wild type
maternal
wild type
Igf2r substitution
maternal
Endogenous Rasgrf1 ♀ Allele
Endogenous Rasgrf1 ♂ Allele
Transgenic Rasgrf1
1147.2 1147.3 1147.4 1147.5 1147.10
Mouse number
– + – + – + – + – + 1kb
HhaI digestion
C
AK006067
AK030170
AK015891 AK029869
10 kbp
DMD Repeat
Rasgrf1
BAC Transgene
wildtype
NO Transgene
R2
Transgenic Rasgrf1
Endogenous

III.6.2 Introduction

Dr. Yoon Jung Park created a *Rasgrf1* BAC transgene containing the entire *Rasgrf1* coding sequence, the DMD and loxP-flanked repeats, and four upstream non-coding RNAs (RC allele). The integration site of the BAC, although not precisely known, is not linked to the endogenous locus on chromosome 9. Using this BAC transgenic RC allele, Dr. Park was able to determine that the sequence contained within this interval is sufficient for imprinting to occur. Using assays originally established for the *Rasgrf1* endogenous locus, she was able to determine the methylation and expression status of the transgene for both maternal and paternal modes of inheritance. When inherited maternally, the RC allele is silent and unmethylated. When inherited paternally, the RC allele is expressed and methylated.

Additionally, Dr. Park observed a *trans*-expression effect at the RC allele transgene. The *trans*-effect is seen in combination with an allele that contains a DMR known as Region 2 from the imprinted and maternally expressed *Igf2r* gene in place of the *Rasgrf1* tandem DNA repeats at the endogenous *Rasgrf1* locus on chromosome 9 (Region 2 allele). When RC is inherited maternally, and in combination with a paternally inherited Region 2 allele, the RC allele is both methylated and expressed (Figure III.6.2).

We wondered whether this *trans*-expression effect could be mediated by small RNAs and, to address this question, used a conditional *Dcr* knockdown allele. The *Dcr* conditional knockdown allele created by Dr. Anders Lindroth contains a GFP reporter under the control of the *CMV* promoter for screening. This cassette is flanked by loxP sites for removal with

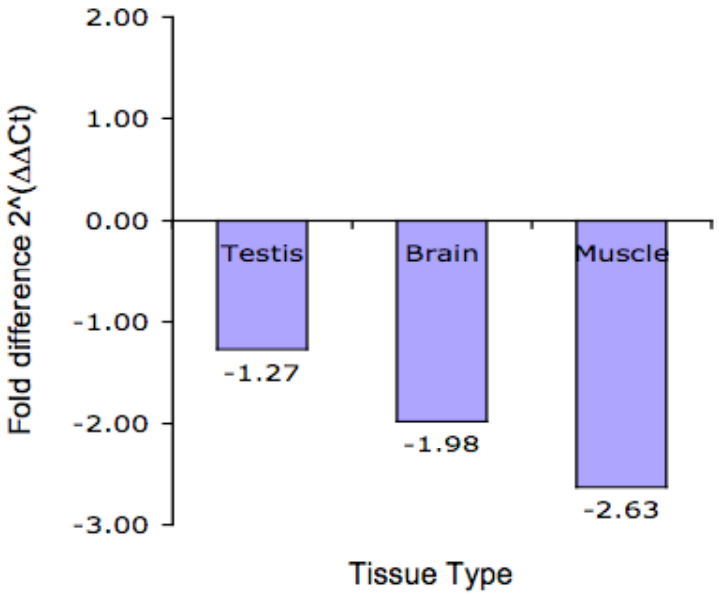


Figure III.6.3. Quantification of *Dcr* expression levels in knockdown animals[16]. Quantitative real-time PCR results quantifying the level of *Dcr* expression in knockdown animals relative to the level of *Dcr* expression in wildtype animals. Data are shown for expression leves in three tissue types, testis, brain, and musle. *Dcr* expression in each tissue type is compared to that in wildtype animals. Both sets of data are standardized to an *Rpl32* internal control amplification. Values reported are fold differences, calculated as 2^(ΔΔCt).

[16] Figure from Dr. Anders Lindroth.

Cre recombinase. Once deleted, an upstream U6 promoter is able to drive expression of a short hairpin RNA with homology to *Dcr*. Although knockdown can never be complete, as *Dcr* is required to produce siRNAs, expression analysis of *Dcr* indicates that knockdown is between 1.27 and 2.53 fold Figure (III.6.3).

To create the desired mice, we first crossed mice carrying the Region 2 allele with mice carrying *Meox2 Cre* recombinase. Then, progeny carrying both the Region 2 allele and *Meox2 Cre* were crossed to mice carrying the unrecombined *Dcr* transgene. A male progeny carrying *Meox2 Cre*, the Region 2 allele, and the unrecombined *Dcr* transgene was then crossed to a female carrying the RC allele. The progeny of this cross were genotyped for *Meox2 Cre*, the Region 2 allele, the **recombined** *Dcr* transgene (which will knock down *Dcr* levels), the RC allele, and the presence of the RC allele repeats, since they are flanked by loxP sites (Figure III.6.4, Figure III.6.5). Progeny with the desired genotype were used to determine the effect of *Dcr* knockdown on the *trans*-expression effect observed between the RC allele and the Region 2 allele.

III.6.3 Materials and methods

III.6.3.1 Timed matings

Females were placed into breeding cages with males overnight and evidence of plugging was checked each morning. At e17.5, female mice were sacrificed and embryos were dissected out.

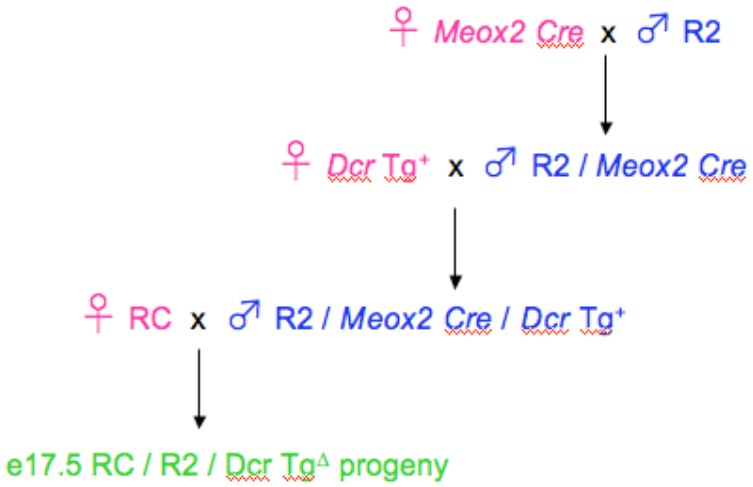


Figure III.6.4. Crossing scheme resulting in RC/*Dcr* TG/Region 2 animals. To create mice carrying the RC BAC transgene, the *Dcr* knockdown transgene, and the Region 2 allele, mice carrying the Region 2 allele were corssed with mice carrying *Meox2 Cre* recombinase. In the next step, progeny carrying both the Region 2 allele and *Meox2 Cre* were crossed to mice carrying the unrecombined *Dcr* transgene. A male progeny inheriting *Meox2 Cre*, the Region 2 allele, and the unrecombined *Dcr* transgene was then crossed to a female carrying the RC allele. The progeny of this cross were genotyped for the unrecombined and the recombined *Dcr* transgene (which will knock down *Dcr* levels), *Meox2 Cre*, the Region2 allele, and the RC allele, as well as for the presence of the RC allele repeats, since they are flanked by loxP sites (Figure III.6.5)

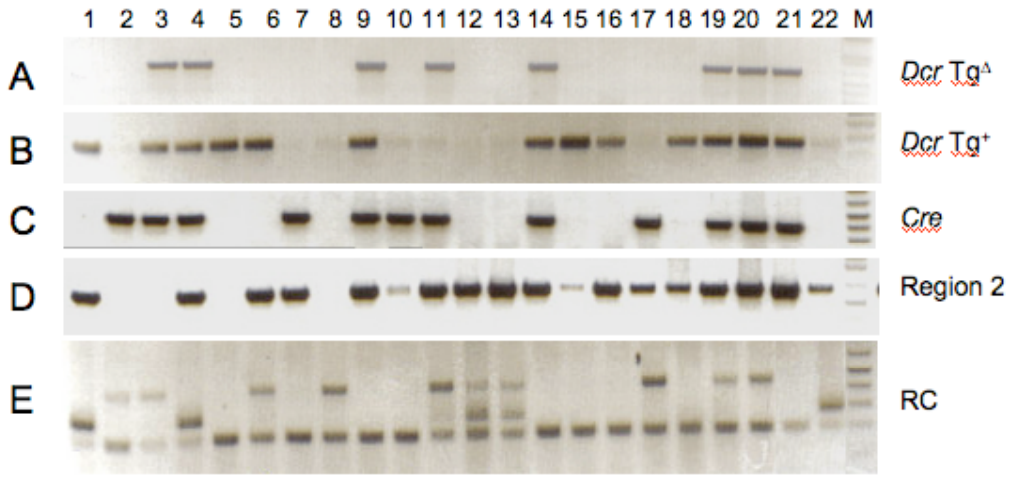


Figure III.6.5 Genotyping for the *Dcr* transgene, *Meox2 Cre,* the Region 2 allele, and RC. Progeny of the crossing scheme in Figure III.6.4 were genotyped for The progeny of this cross were genotyped for the presence (B) and recombimation of (A) the *Dcr* transgene, *Meox2 Cre* (C), the Region 2 allele (D), and the RC allele (E). The RC allele genotype also provided information on the status of the RC allele's loxP-flanked DNA repeats. The bottom band is specific to the wildtype endogenous repeats and serves as an internal control. The middle band is indicative of unrecombined loxP-flanked repeats present on the RC allele, and the top band indicated that the loxP-flanked repeats of the RC allele have been deleted by *Meox2 Cre* recombinase.

III.6.3.2 DNA isolation

During embryo dissection, a tail snip and hindlimb were reserved and lysed with filter-sterilized Laird's lysis buffer (50 mM pH 8.0 Tris, 100 mM EDTA, 100mM NaCl, 1% SDS, and 100ug/mL proteinase K) overnight at 55C with shaking at 220rpm. DNA was then prepared by isoproanol precipitation for genotyping.

III.6.3.3 Genotyping

To detect the presence of *Cre*, *Cre*1 (5' – TGA TGA GGT TCG CAA GAA CC – 3') and *Cre*2 (5' – CCA TGA GTG AAC GAA CCT GG – 3') were used under the conditions 95C for 10sec, 58C for 20sec, 72C for 30sec for 37 cycles. To determine the extent of repeat-deletion at the RC locus, PDS16 (5' – GCA CTT CGC TAC CGT TTC GC – 3'), PDS18 (5' – TTT CTG CCA TCA TCC CAG CC – 3'), and PDS17 (5' – TGT CCT CCA CCC CTC CAC C– 3') were used under the conditions 94C for 10sec, 58C for 20sec, 72C for 30sec for 37 cycles. To detect the presence of the unrecombined and inactive *Dicer* knockdown transgene (*Dcr* Tg+), PDS194 (5' – CAT GTC ACA AAA GGA AAC TCA CCC – 3') and PDS195 (5' – GGC TAT GAA CTA ATG ACC CCG TAA – 3') were used under the conditions 95C for 10sec, 58C for 20sec, 72C for 30sec for 35 cycles. To detect the presence of the recombined and active *Dicer* knockdown transgene (*Dcr* Tg Δ), PDS71 (5' – CAA ACA CAG TGC ACA CCA CGC – 3') were used under the conditions 94C for 10sec, 58C for 20sec, 72C for 30sec for 37 cycles. To detect the RC BAC transgene, PDS16 (5'- GCA CTT CGC TAC CGT TTC GC -3') and pYJC6R2 (5'- CCT GCA GGT CGA CAT AAC TTC -3') were used under the conditions 94C for 10sec, 61C for 20sec, 72C for 50sec for 40 cycles. To detect the Region 2 allele, Lu920F

and Lu1520R were used (sequence not available). However, primers PDS16 (5' – GCA CTT CGC TAC CGT TTC GC – 3') and PDS781 (5' – CCT CTA GAG TCG ACA TAA CTT CG– 3') can also be used under the conditions 94C for 30sec, 64C for 1min, 72C for 30sec for 40 cycles.

III.6.3.4 RNA and cDNA preparation

RNA was prepared from neonatal brains by homogenization at 18,000rpm for 45 seconds in 2mL GTC lysis buffer (20mls GTC solution (4M guanine thiocyanate, 25mM sodium citrate, 5% sarcosyl), 3M β-mercaptoethanol, 20mls acidic phenol, 2mls 2M sodium acetate). After homogenization, 400ul chloroform was added, lightly vortexed and centrifuged at 10,000g for 15 minutes at 4C. The aqueous phase was transferred to a new tube, precipitated with 0.7 volumes 100% isopropanol for 20 minutes at -80C, and spun at 10,000g for 15 minutes at 4C. The pellet was rinsed with 70% ethanol, air dried, and resuspended in 120ul DEPC-TE.

cDNA was prepared from 5ug RNA per sample. Samples were DNaseI treated (2.5ul 10X DNaseI buffer, 2.5ul DNaseI, water to 20ul) for 45 minutes at room temperature and reaction was quenched with 1ul of 25mM EDTA at 65C for 10 minutes. The DNaseI treated samples were split in half and one half was frozen as a –RT sample. 0.7ul of random hexamers, 1.5ul 10mM dNTPs, and 0.8ul DEPC H2O was added to the second half of the reaction and was incubated at 65C for 5 minutes followed by 2 minutes on ice. Then, 6ul of 5X RT buffer, 3ul of 0.1M DTT, 1.5ul RNaseOUT, and 4ul of DEPC H20 was added at 42C for 2 minutes, followed by a spike of 0.5ul SuperscriptII reverse transcriptase. The synthesis reaction was carried out at 42C for 1hour

followed by 70C for 15 minutes and held at 4C.

III.6.3.5 Expression analysis

PCR was carried out on cDNA samples using the primers PDS245 (5' – GGC TCA TGA TGA ATG CCT TT – 3') and PDS246 (5' – TAC AGA AGC TTG GCG TTG TG – 3') under cycling conditions 95C for 20 seconds, 58C for 30 seconds, 72C for 50 seconds for 38 cycles. After PCR, the resulting product was digested overnight with AciI. The presence of a restriction fragment length polymorphism allows detection of both endogenous and RC1 *Rasgrf1* expression after digestion with AciI.

III.6.3.6 HhaI methylation analysis

500ng of DNA was added to a 1.5ml tube labeled "uncut". A HhaI restriction digestion master mix was made up to a total of 78ul, but no HhaI was added. The tube was then split in half and 38ul were added to a second tube. 1ul of HhaI was added to the second tube, which was labeled "cut". Both tubes were incubated at 37C overnight. In the morning, 0.5ul of HhaI was spiked in and the digestion was allowed to proceed all day at 37C. That evening, 0.5ul more HhaI was spiked and the digestion was allowed to proceed overnight at 37C. In the morning, both tubes were incubated at 65C for 20 minutes to inactivate the HhaI. Then, a digestion control PCR was set up using primers PGKF2 (5' – CTT TGC TCC TTC GCT TTC TG - 3') and PGKR2 (5' – ACG TCC AGC TTG TCC AAA GT - 3') under conditions 94C for 30 seconds, 62C for 30 seconds, 72C for 50 seconds for 40 cycles. An amplification control PCR was set up using primers Rpl32F (5' –CAT GCA CAC AAG CCA TCT ACT CA - 3') and Rpl32R (5' – TGC TCA CAA TGT GTC

CTC TAA GAA C - 3') under conditions 94C for 30 seconds, 62C for 30 seconds, 72C for 50 seconds for 40 cycles. A methylation analysis PCR was set up using primers PDS16 (5' – GCA CTT GCG TAC CGT TTC GC - 3') and PDS17 (5' – TGT CCT CCA CCC CTC CAC C - 3') under conditions 94C for 30 seconds, 62C for 30 seconds, 72C for 50 seconds for 40 cycles. All three PCR reactions were run out on a 1.5% agarose gel.

III.6.4 Results

Embryonic day 17.5 pups were dissected out and a portion of the embryo was genotyped for the presence of the RC transgene, the Region 2 allele, and the recombined version of the *Dicer* knockdown transgene (*Dcr* TgΔ), as well as for the absence of *Cre* and the unrecombined version of the *Dicer* knockdown transgene (*Dcr* Tg+) (Figure III.6.5). Out of 64 pups genotyped, 8 had the correct combination of alleles, except that all were *Cre* positive and, therefore, had repeat-deleted versions of the RC BAC transgene. As a control for the effect of deletion of the RC repeats, I identified a *Dcr* transgene-less mouse carrying both a repeat-deleted version of RC and Region 2. A PCR assay was used to determine the extent of repeat deletion, ensuring that this animal was not mosaic for deletion of the RC repeats. Additionally, all but two of the samples tested for *trans* expression of the RC allele showed complete deletion of the RC repeats (samples 2 and 3 in Figure III.6.6 are mosaic for deletion of the RC repeats) Then, I prepared cDNA from e17.5 brain and performed expression PCR using primers specific to *Rasgrf1*. Through the use of a restriction fragment length polymorphism, these primers will allow detection of both endogenous and RC1 *Rasgrf1* expression after

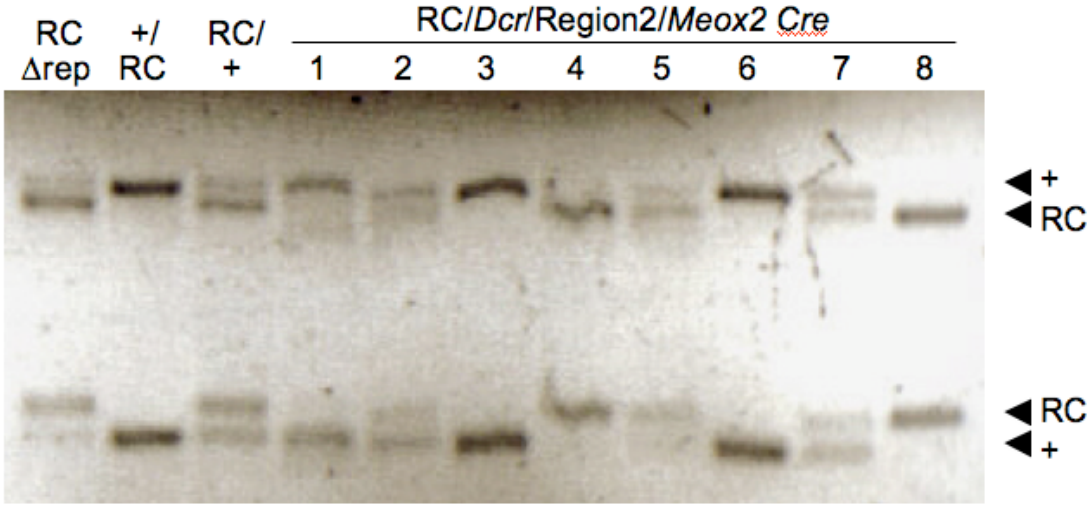


Figure III.6.6. Expression analysis of RC/*Dcr* knockdown/Region 2/*Meox2 Cre* animals. RT-RCR was done on e17.5 brain samples collected from RC/*Dcr* knockdown/Region 2/*Meox2 Cre* progeny, as well as from +/RC and RC/+ controls. The lane labeled RC Δrep is an additional control for the effect of the deletion of the RC loxP-flanked repeats. This animal is positive for RC, Region 2, and *Meox2 Cre* and the RC DNA repeats have been deleted. Therefore, any expression effects must be due to the deletion of the repeats and not to a knockdown of *Dcr.* Expression in this animal is from both the endogenous *Rasgrf1*, as well as the RC BAC transgene, indicating that the RC repeats are not needed for the trans effect to occur. Samples labeled 1-8 are the RC/*Dcr* knockdown/Region 2/*Meox2 Cre* progeny. Expression is variable between endogenous *Rasgrf1* only (bands marked +), RC only (bands marked RC), and expression from both the endogenous *Rasgrf1* and the RC BAC transgene.

digestion with AciI. After PCR, I did allele-specific restriction digest on expression on the resulting products (Figure III.6.6).

The expression results indicate variable expression patterns. In some cases, expression is from the endogenous *Rasgrf1* allele. In other cases, expression is from the RC allele, or from both alleles. Analysis of the repeat-deleted RC animal indicates that deletion of the RC repeats does not affect trans-expression of RC. Therefore, any expression changes seen are due to the knockdown of *Dicer* and not to the deletion of the RC repeats. After seeingthe variable results from the expression analysis, I did a methylation analysis on the same samples to quantify the methylation levels at the RC transgene allele. The methylation analysis revealed lack of expression from the RC allele roughly corresponded with lack of methylation at the RC DMD, although a notable exception is that samples 6 and 8 show the same methylation profile, but opposite expression patterns (Figure III.6.6 and Figure III.6.7).

III.6.5 Discussion

Although there does seem to be some link between knocking down *Dicer* and disruption of the RC trans-expression effect, the results seen so far have been variable. It is possible that the different samples tested demonstrate different degrees of *Dicer* knockdown. To test this possibility, levels of *Dicer* expression could be tested using quantitative real-time RT-PCR.

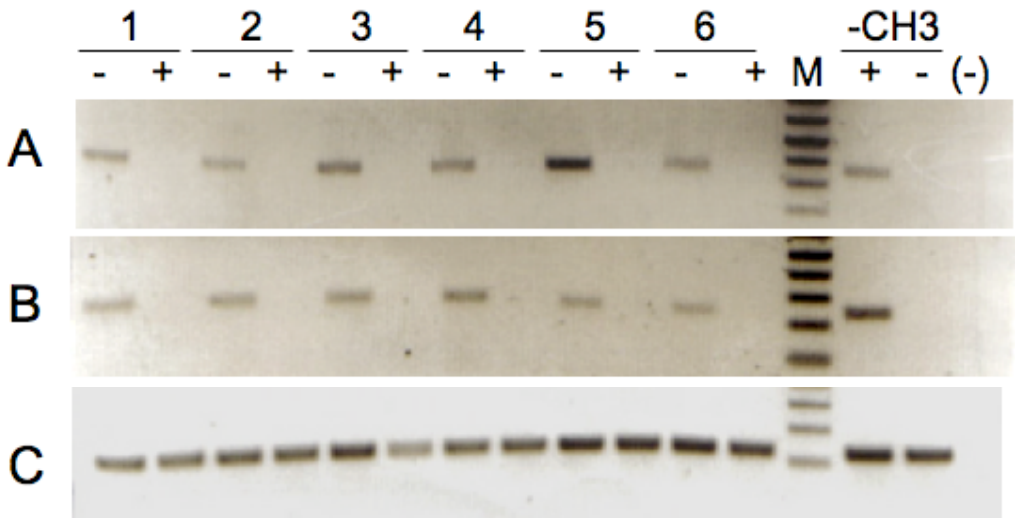


Figure III.6.7. Methylation analysis of RC/*Dcr* knockdown/Region 2/*Meox2 Cre* progeny used in expression analysis. HhaI digestion analysis of methylation levels at the RC BAC transgene DMD of RC/*Dcr* knockdown/Region 2/*Meox2 Cre* progeny used in the expression analysis in Figure III.6.6. (A) PCR assay specific to the RC BAC transgene DMD. Presence of the bands in the cut lane (+) indicates methylation at the DMD. (B) PGK digestion control. Presence of a PCR band in the cut lane (+) indicates incomplete HhaI digestion. (C) Rpl32 amplification control. Also included in each assay is PCR analysis of the uncut DNA (-), which serves as a positive control for each DNA sample. A known unmethylated control was also included (-CH3), as was a water control ((-)).

The methylation analysis suggests that trans-expression of the RC allele is dependent on the level of methylation at the RC1 allele. Of course, the sensitivity of this method of methylation analysis is much less ideal than that provided by other methods. A good first step would be to confirm and expand upon the results of the HhaI digest for methylation levels at the RC DMD using Sequenom MASSarray. It would also be interesting to check the methylation level of the endogenous maternal allele by both HhaI digestion and Sequenom MASSarray analysis. The methylation results, again, could be linked to the level of *Dicer* knockdown. It is known that knockdown of *Dicer* leads to reduced levels of *Dnmt3b*, which could lead to the reduced levels of DNA methylation at the RC DMD. To explore the role of *Dnmt3b* (and other *Dnmt* genes) expression levels in the trans effect at RC, the expression levels of all three *Dnmt* genes should be examined using quantitative real-time RT-PCR. Of course, there is also the possibility that DNA methylation is controlled directly by *Dicer* through small RNAs. As discussed above (see section III.3), our lab has identified a piRNA with homology to the *Rasgrf1* repeats, and it will be interesting to see whether DNA methylation in the *Rasgrf1* system is controlled by *Dnmt* levels, by small RNAs, or by a combination of these two mechanisms.

III.7 NECESSITY OF THE MATERNAL REPEATS FOR A *TRANS*-EFFECT AT *RASGRF1*

III.7.1 Abstract

Rasgrf1 is a paternally expressed imprinted gene located on mouse chromosome 9. Imprinted expression of *Rasgrf1* is controlled by a DMD and a series of tandem DNA repeats (Yoon 2002). Replacing the *Rasgrf1* repeats with the Region 2 differentially methylated region (Region 2 allele) from the imprinted and maternally expressed gene, *Igf2r*, allows for DNA methylation and expression of the paternal allele of *Rasgrf1* (Herman 2003). Paternal transmission of the Region 2 allele also results in both methylation and expression, in *trans,* of the normally unmethylated and silent wildtype maternal allele. I am investigating whether the presence of the tandem DNA repeats on the maternal allele is necessary for this trans effect to occur.

III.7.1 Introduction

Rasgrf1 is a paternally expressed imprinted gene on mouse chromosome 9. Expression of *Rasgrf1* is controlled by a binary switch comprised of a differentially metnylated domain (DMD) and a series of DNA repeats, located 30kb upstream of the *Rasgrf1* promoter. The DNA repeats consist of 40 copies of a 41bp repeat unit and provide a positive signal for placement of DNA methylation at the DMD, which leads to expression of *Rasgrf1* from the paternal allele in neonatal brain (Yoon 2002). The DNA repeats were shown to be necessary for the establishment of DNA methylation in the male germline (Yoon 2002) and for the maintenance of DNA methylation

between e0.0 and e5.5 (Holmes 2006). However, after e5.5 (when genomic methylation reprogramming has finished) the DNA repeats are dispensable.

Despite the importance of the DNA repeats for the establishment and maintenance of proper DNA methylation, replacing the *Rasgrf1* repeats with the Region 2 differentially methylated region (Region 2 allele) from the imprinted and maternally expressed gene, *Igf2r*, allows for DNA methylation and expression of the paternal allele of *Rasgrf1* (Herman 2003). Interestingly, paternal transmission of the Region 2 allele also results in both methylation and expression, in *trans,* of the normally unmethylated and silent wildtype maternal allele. Additionally, the activated wildtype maternal allele maintains its active state in subsequent generations, even in the absence of a Region 2 paternal allele (Herman 2003).

Based on preliminary evidence from 3C and FISH analysis (see Appendix sections III.4 and III.5), we had reason to believe that an inter-chromosomal interaction may be occurring in the *Rasgrf1* imprinted region. Since the DNA repeats are important for both the establishment and the maintenance of DNA methylation at *Rasgrf1* under wildtype conditions, we wondered whether the repeats were somehow involved in mediating DNA methylation on the maternal allele in the presence of a Region 2 paternal allele. To investigate this possibility, I set up crosses between a female homozygous for a deletion of the *Rasgrf1* tandem DNA repeats and a male heterozygous for the Region 2 allele and examined their progeny for evidence of DNA methylation on the maternal allele.

III.7.2 Materials and Methods

III.7.2.1 Genotyping

To detect the Region 2 allele, primers PDS16 (5' – GCA CTT CGC TAC CGT TTC GC – 3') and PDS781 (5' – CCT CTA GAG TCG ACA TAA CTT CG– 3') can be used under the conditions 94C for 30sec, 64C for 1min, 72C for 30sec for 40 cycles.

III.7.2.2 DNA extraction

Tissue samples were lysed overnight in filter-sterilized Laird's lysis buffer (50 mM pH 8.0 Tris, 100 mM EDTA, 100mM NaCl, 1% SDS, and 100ug/mL proteinase K) at 55C with shaking at 220rpm. DNA was extracted by isopropanol precipitation.

III.7.2.3 Methylation analysis

500ng of DNA was added to a 1.5ml tube labeled "uncut". A HhaI restriction digestion master mix was made up to a total of 78ul, but no HhaI was added. The tube was then split in half and 38ul were added to a second tube. 1ul of HhaI was added to this tube, labeled "cut". Both tubes were incubated at 37C overnight. In the morning, 0.5ul of HhaI was spiked in and the digestion was allowed to proceed all day at 37C. That evening, 0.5ul more HhaI was spiked and the digestion was allowed to proceed overnight at 37C. In the morning, both tubes were incubated at 65C for 20 minutes to inactivate the HhaI. Then, a digestion control PCR was set up using primers PGKF2 (5'

– CTT TGC TCC TTC GCT TTC TG - 3') and PGKR2 (5' – ACG TCC AGC TTG TCC AAA GT - 3') under conditions 94C for 30 seconds, 62C for 30 seconds, 72C for 50 seconds for 40 cycles. An amplification control PCR was set up using primers Rpl32F (5' – CAT GCA CAC AAG CCA TCT ACT CA - 3') and Rpl32R (5' – TGC TCA CAA TGT GTC CTC TAA GAA C - 3') under conditions 94C for 30 seconds, 62C for 30 seconds, 72C for 50 seconds for 40 cycles. Primers used for the methylation analysis were PDS16 (5' – GCA CTT CGC TAC CGT TTC GC – 3') and PDS17 (5' – TGT CCT CCA CCC CTC CAC C– 3'). These were used under the conditions 94C for 10sec, 58C for 20sec, 72C for 30sec for 37 cycles. All three PCR reactions were run out on a 1.5% agarose gel

III.7.3 Results

Progeny of a cross between female mice homozygous for a deletion of the *Rasgrf1* tandem DNA repeats and a male heterozygous for the Region 2 allele were genotyped for the presence of the Region 2 allele. Tail DNA from pups inheriting a paternal Region 2 allele was subjected to HhaI methylation analysis. PCR using either 43 or 55 cycles showed no evidence of *trans* DNA methylation at the maternal *Rasgrf1* DMD (Figure III.7.1).

III.7.4 Discussion

DNA from mice inheriting a *Rasgrf1* allele lacking the tandem DNA repeats maternally and a Region 2 allele paternally was subjected to HhaI

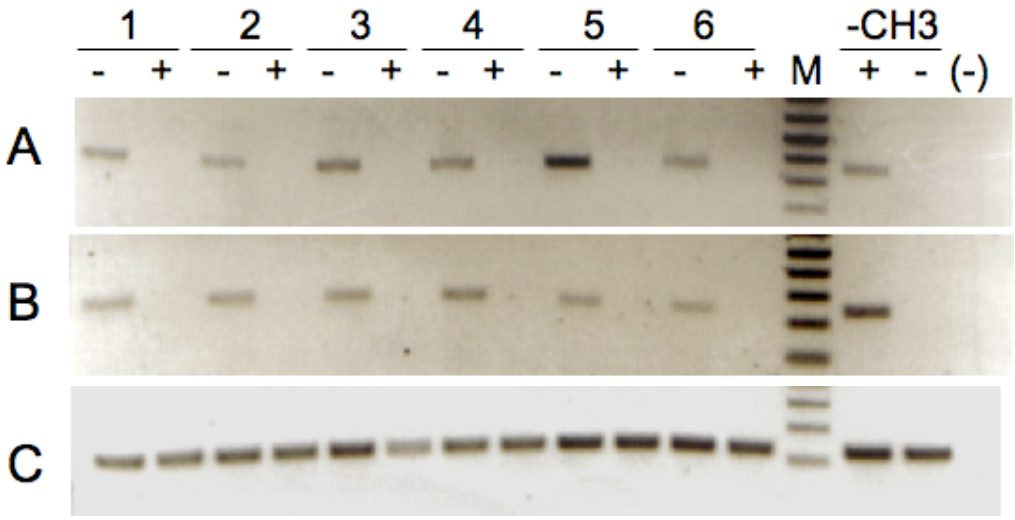


Figure III.7.1. Analysis of *trans*-methylation at the maternal DMD. HhaI digestion analysis of methylation levels at the maternal DMD of progeny inheriting a paternal Region 2 allele. (A) PCR assay specific to the maternal DMD. Presence of the bands in the cut lane (+) indicates methylation at the DMD. (B) PGK digestion control. Presence of a PCR band in the cut lane (C) indicates incomplete HhaI digestion. (C) Rpl32 amplification control. Also included in each assay is PCR analysis of the uncut DNA (-), which serves as a positive control for each DNA sample. A known unmethylated control was also included (-CH3) as was a water control ((-)).

methylation analysis. PCR using either 43 or 55 cycles showed no evidence of DNA methylation at the maternal *Rasgrf1* DMD (Figure III.7.1). This result suggests that the maternal tandem DNA repeats are necessary for the *trans*-methylation effect caused by paternal inheritance of the Region 2 allele.

However, there is a distinct possibility that DNA methylation is still occurring on the maternal allele even in the presence of a repeat deletion. The *trans*-methylation that occurs on the maternal allele does so at very low levels. It is possible that the PCR assay used to detect methylation at the *Rasgrf1* DMD is not sensitive enough to report the very low levels of DNA methylation that occur on the maternal allele. Therefore, it is important to perform additional tests for PCR sensitivity. One possibility is that the same primers used to detect DNA methylation at the repeat-deleted DMD could be used on progeny inheriting a wildtype maternal DMD and a paternal Region 2 allele. However, this PCR product would be about 2kb in size and may be difficult to amplify in the first place. Any small amount of *trans*-methylation therefore may not be detected. Another possibility is to perform radioactive PCR on HhaI digested DNAs from progeny inheriting a repeat-deleted maternal DMD and a paternal Region 2 allele. Until additional tests are done, no conclusions can be drawn from these results.

REFERENCES

Alleman M., Sidorenko L., McGinnis K., Seshadri V., Dorweiler J.E., White J., Sikkink K., and Chandler V.L. (2006). An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* 442, 295-298.

Anway M.D., Cupp A.S., Uzumcu M., and Skinner M.K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308, 1466-1469.

Aravin A.A., Sachidanandam R., Girard A., Fejes-Toth K., Hannon G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744-7.

Aravin A.A., Sachidanandam R., Bourc'his D., Schaefer C., Pezic D., Toth K.F., Bestor T., Hannon G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* 31, 785-99.

Aufsatz W., Mette M.F., van der Winden J., Matzke A.J., Matzke M. (2002). RNA-directed DNA methylation in Arabidopsis. *Proc Natl Acad Sci U S A.* 99, 16499-506.

Babak T., Deveale B., Armour C., Raymond C., Cleary M.A., van der Kooy D., Johnson J.M., Lim L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol.* 18, 1735-41.

Barlow D.P. (1995). Gametic imprinting in mammals. *Science* 270, 1610–1613.

Bartolomei M. and Tilghman S. (1997). Genomic imprinting in mammals. *Annu. Rev. Genet*. 31, 493-525.

Barton S.C., Surani M.A., and Norris M.L. (1984). Role of paternal and maternal genomes in mouse development. *Nature* 311, 374-376.

Bastepe M., Fröhlich L.F., Hendy G.N., Indridason O.S., Josse R.G., Koshiyama H., Körkkö J., Nakamoto J.M., Rosenbloom A.L., Slyper A.H., Sugimoto T., Tsatsoulis A., Crawford J.D., Jüppner H. (2003) Autosomal dominant pseudohypoparathyroidism type Ib is associated with a heterozygous microdeletion that likely disrupts a putative imprinting control element of GNAS. *J. Clin. Invest.* 112: 1255–1263.

Bell A.C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482-485.

Benetti R., Gonzalo S., Jaco I., Muñoz P., Gonzalez S., Schoeftner S., Murchison E., Andl T., Chen T., Klatt P., Li E., Serrano M., Millar S., Hannon G., Blasco M.A. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via *Rbl2*-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol*.15, 268-79.

Bernstein B.E., Mikkelsen T.S., Xie X., Kamal M., Huebert D.J., Cuff J., Fry B., Meissner A., Wernig M., Plath K., Jaenisch R., Wagschal A., Feil R., Schreiber S.L., Lander E.S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6-21.

Birger Y., Shemer R., Perk J., and Razin A. (1999). The imprinting box of the mouse *Igf2r* gene. *Nature* 397, 84-88.

Boily G., Seifert E.L., Bevilacqua L., He X.H., Sabourin G., Estey C., Moffat C., Crawford S., Saliba S., Jardine K., Xuan J., Evans M., Harper M.E., McBurney M.W. (2008). *SirT1* regulates energy metabolism and response to caloric restriction in mice. *PLoS ONE* 3, e1759.

Bourc'his, D. and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking *Dnmt3L*. *Nature* 431, 96-99.

Boyes J., Bird A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64, 1123-34.

Brennecke J., Aravin A.A., Stark A., Dus M., Kellis M., Sachidanandam R., Hannon G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089-103.

Brown C.J., Ballabio A., Rupert J.L., Lafreniere R.G., Grompe M., Tonlorenzi R., Willard H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349, 38-44.

Buiting K., Saitoh S., Gross S., Dittrich B., Schwartz S., Nicholls R.D., Horsthemke B. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet*. 9, 395-400.

Burge S., Parkinson G.N., Hazel P., Todd A.K., and Neidle S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucl. Acids Res.* 34, 5402-5415.

Cai X. and Cullen B.R. (2007). The imprinted *H19* noncoding RNA is a primary microRNA precursor. *RNA.* 13, 313-6.

Caretti G., Di Padova M., Micales B., Lyons G.E., and Sartorelli V. (2004). The Polycomb *Ezh2* methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev.* 18, 2627-2638.

Carmell M.A., Girard A., van de Kant H.J., Bourc'his D., Bestor T.H., de Rooij D.G., Hannon G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell*. 12, 503-14.

Carrozza M.J., Li B., Florens L., Suganuma T., Swanson S.K., Lee K.K., Shia

W.J., Anderson S., Yates J., Washburn M.P., Workman J.L. (2005). Histone H3 methylation by *Set2* directs deacetylation of coding regions by *Rpd3S* to suppress spurious intragenic transcription. *Cell* 123, 581-592.

Cattanach B.M. and Kirk M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315, 496-498.

Cattanach B.M., Barr J.A., Evans E.P., Burtenshaw M., Beechey C.V., Leff S.E., Brannan C.I., Copeland N.G., Jenkins N.A., Jones J. (1992). A candidate mouse model for Prader-Willi syndrome which shows an absence of *Snrpn* expression. *Nat. Genet*. 2, 270-4.

Cen H., Papageorge A.G., Vass W.C., Zhang K., and Lowry D.R.. (1993). Regulated and constitutive activity by CDC25^{Mm} (GRF), a *Ras*-specific exchange factor. *Mol. Cell Biol.* 13, 7718-7724.

Chan S.W., Zilberman D., Xie Z., Johansen L.K., Carrington J.C., and Jacobsen S.E. (2004). RNA silencing genes control *de novo* DNA methylation. *Science* 303, 1336.

Chan S.W., Henderson I.R., and Jacobsen S.E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev.* 6, 351-360.

Chan S.W., Zhang X., Bernatavichute Y.V., and Jacobsen S.E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* 4, e363.

Chao W., Huynh K.D., Spencer R.J., Davidow L.S., and Lee J.T. (2002). CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science* 295, 345-347.

Cho D.H., Thienes C.P., Mahoney S.E., Analau E., Filippova G.N., and Tapscott S.J. (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* 20, 483-489.

Cohen D.E., Davidow L.S., Erwin J.A., Xu N., Warshawsky D., Lee J.T. (2007). The *DXPas34* repeat regulates random and imprinted X inactivation. *Dev. Cell*. 12, 57-71.

Crespi, B. (2008). Genomic imprinting in the development and evolution of psychotic spectrum conditions. *Biol. Rev. Camb. Philos. Soc*. 83, 441-93.

Danielian P.S., Echelard Y., Vassileva G., McMahon A.P. (1997). A 5.5-kb enhancer is both necessary and sufficient for regulation of *Wnt-1* transcription in vivo. *Dev. Biol.* 192, 300-9.

da Rocha S.T., Tevendale M., Knowles E., Takada S., Watkins M., Ferguson-Smith A.C. (2007). Restricted co-expression of *Dlk1* and the reciprocally imprinted non-coding RNA, *Gtl2*: implications for cis-acting control. *Dev. Biol*. 306:810-23.

de la Puente A., Hall J., Wu Y.Z., Leone G., Peters J., Yoon B.J., Soloway P.,

and Plass C. (2002). Structural characterization of *Rasgrf1* and a novel linked imprinted locus. *Gene* 291, 287-297.

Delaval K., Govin J., Cerqueira F., Rousseaux S., Khochbin S., and Feil R. (2007). Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* 26, 720-729.

Delaval K., Wagschal A., and Feil R. (2006). Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *Bioessays* 28, 453-459.

Dobosy, J.R. and Selker E.U. (2001). Emerging connections between DNA methylation and histone acetylation. *Cell. Mol. Life Sci.* 58, 721-727.

Donohoe M.E., Zhang L.F., Xu N., Shi Y., and Lee J.T. (2007). Identification of a Ctcf cofactor, *Yy1*, for the X chromosome binary switch. *Mol. Cell* 25, 43-56.

Donohoe M.E., Zhang X., McGinnis L., Biggers J., Li E., and Shi Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell Biol.* 19, 7237-7244.

Eads C.A., Nickel A.E., and Laird P.W. (2002). Complete genetic suppression of polyp formation and reduction of CpG-island hypermethylation in *Apc*$^{Min/+}$*Dnmt1*-hypomorphic mice. *Cancer Res.* 62, 1296-1299.

Ebbs M.L. and Bender, J. (2006). Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell* 18, 1166-

1176.

Echelard Y., Vassileva G., and McMahon A.P. (1994). *Cis*-acting regulatory sequences governing *Wnt1* expression in the developing mouse CNS. *Development* 120, 2212-2224.

Edwards C.A. and Ferguson-Smith A.C. (2007). Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* 19, 281-289.

Engel, E. (1980). A new genetic concept: uniparental disomy and its potential effect, isodisomy. *Am. J. Med. Genet*. 6, 137-143.

Erhardt S., Su I.H., Schneider R., Barton S., Bannister A.J., Perez-Burgos L., Jenuwein T., Kouzarides T., Tarakhovsky A., and Surani M.A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* 130, 4235-4248.

Espada J., Ballestar E., Fraga M.F., Villar-Garea A., Juarranz A., Stockert J.C., Robertson K.D., Fuks F., and Esteller M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *J. Bio.l Chem.* 279, 37175-37184.

Esteve P.O., Chin H.G., Smallwood A., Feehery G.R., Gangisetty O., Karpf A.R., Carey M.F., and Pradhan S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* 20, 3089-3103.

Fedoriw A.M., Stein P., Svoboda P., Schultz R.M., and Bartolomei M.S. (2004). Transgenic RNAi reveals essential function for CTCF in *H19* gene imprinting. *Science* 303, 238-240.

Feinberg A.P. and Vogelstein B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89-92.

Ferguson-Smith A.C., Sasaki H., Cattanach B.M., Surani M.A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 362, 751-5.

Filippova G.N., Fagerlie S., Klenova E.M., Myers C., Dehner Y., Goodwin G., Neiman P.E., Collins S.J., and Lobanenkov V.L. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian *c-myc* oncogenes. *Mol. Cell Biol.* 1996; 16: 2802-2813.

Goll M.G. and Bestor T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. of Biochem.* 74, 481-514.

Greally J.M. (2002). Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl. Acad. Sci. U S A*99, 327-332.

Guarente L. and Picard, F. (2005). Calorie restriction--the SIR2 connection.

Cell 120, 473-482.

Haim N. Methylation: Gene expression at the right place and right time. *The Scientist* 1999; 13: 21.

Hajkova P., Erhardt S., Lane N., Haaf T., El-Maarri O., Reik W., Walter J., and Surani M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15-23.

Hardin C.C., Corregan M., Brown B.A., and Frederick L.N. (1993). Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. *Biochemistry* 32, 5870-5880.

Hata K., Okano M., Lei H., Li E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983-93.

Hatada I., Sugama T., Mukai T. (1993). A new imprinted gene cloned by a methylation-sensitive genome scanning method. *Nucl. Acids Res*. 21, 5577-82.

Hayashizaki Y., Shibata H., Hirotsune S., Sugino H., Okazaki Y., Sasaki N., Hirose K., Imoto H., Okuizumi H,. Muramatsu M*.,* Komatsubara H., Shiroishi T., Moriwaki K., Katsuki M., Hatano N., Sasaki H., Ueda T., Mise N., Takagi N., Plass C., and Chapman V.M. (1994). Identification of an imprinted *U2af* binding protein related sequence on mouse chromosome 11 using the RLGS

method. *Nat. Genet.* 6, 33-40.

Heintzman N.D. Stuart R.K., Hon G., Fu Y., Ching C.W., Hawkins R.D., Barrera L.O., Van Calcar S., Qu C., Ching K.A., Wang W., Weng Z., Green R.D., Crawford G.E., Ren B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311-8.

Herman H., Lu M., Anggraini M., Sikora A., Chang Y., Yoon B.J., and Soloway P.D. (2003). Trans allele methylation and paramutation-like effects in mice. *Nat. Genet.* 34, 199-202.

Hernández-Muñoz I., Taghavi P., Kuijl C., Neefjes J., van Lohuizen M. (2005). Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol Cell Biol.* 25, 11047-58.

Hiby S.E., Lough M., Keverne E.B., Surani M.A., Loke Y.W., and King A. (2001) Paternal monoallelic expression of PEG3 in the human placenta. *Hum. Mol. Genet*. 10:1093-100.

Hoki Y., Kimura N., Kanbayashi M., Amakawa Y., Ohhata T., Sasaki H., Sado T. (2009). A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development*. 136, 139-46.

Holliday R. (1990). DNA methylation and epigenetic inheritance. *Philos. Trans.*

R. Soc. Lond. Biol. Sci. 326, 329-38.

Holmes R., Chang Y., and Soloway P.D. (2006). Timing and sequence requirements defined for embryonic maintenance of imprinted DNA methylation at *Rasgrf1*. *Mol. Cell Biol.* 26, 9564-9570.

Howell C.Y., Bestor T.H., Ding F., Latham K.E., Mertineit C., Trasler J.M., and Chaillet J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* 104, 829-838.

Huppert J.L. and Balasubramanian S. (2005). Prevalence of quadruplexes in the human genome. *Nucl. Acids Res*. 33, 2908-2916.

Ishihara K., Oshimura M., and Nakao M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* 23, 733-742.

Jackson J.P., Lindroth A.M., Cao X., and Jacobsen S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556-560.

Jaenisch R. and Bird A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33 Supplement, 245-254.

Jiang Y.H., Bressler J., and Beaudet A.L. Epigenetics and human disease. *Annu. Rev. Genomics Hum. Genet.* 5, 479-510 (2004).

Johnson L., Cao X., and Jacobsen S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* 12, 1360-1367.

Joshi A.A. and Struhl K. (2005). *Eaf3* chromodomain interaction with methylated H3-K36 links histone deacetylation to *Pol II* elongation. *Mol. Cell* 20, 971-978.

Kaati G., Bygren L.O., Pembrey M., and Sjostrom M. (2007). Transgenerational response to nutrition, early life circumstances and longevity. *Eur. J. Human Genet.* 15, 784-790.

Kaneda M., Okano M., Hata K., Sado T., Tsujimoto N., Li E., and Sasaki H. (2004). Essential role for de novo DNA methyltransferase *Dnmt3a* in paternal and maternal imprinting. *Nature* 429, 900-903.

Kaneko-Ishino T., Kuroiwa Y., Miyoshi N., Kohda T., Suzuki R., Yokoyama M., Viville S., Barton S.C., Ishino F., Surani M.A. (1995). *Peg1/Mest* imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat. Genet.* 11, 52-9.

Kangaspeska S., Stride B., Metivier R., Polycarpou-Schwarz M., Ibberson D., Carmouche R.P., Benes V., Gannon F., Reid G. (2008). Transient cyclical methylation of promoter DNA. *Nature* 452, 112–115.

Kantor B., Makedonski K., Green-Finberg Y., Shemer R., and Razin A. (2004). Control elements within the PWS/AS imprinting box and their function in the imprinting process. *Hum. Mol. Genet.* 13, 751-762.

Kato Y., Kaneda M., Hata K., Kumaki K., Hisano M., Kohara Y., Okano M., Li E., Nozaki M., and Sasaki H. (2007). Role of the *Dnmt3* family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272-2280.

Kato Y. and Sasaki H. (2005). Imprinting and looping: epigenetic marks control interactions between regulatory elements. *Bioessays*. 27, 1-4.

Ke X., Thomas N.S., Robinson D.O., and Collins A. (2002). The distinguishing sequence characteristics of mouse imprinted genes. *Mamm. Genome* 13, 639-645.

Keogh M.C., Kurdistani S.K., Morris S.A., Ahn S.H., Podolny V., Collins S.R., Schuldiner M., Chin K., Punna T., Thompson N.J., Boone C., Emili A., Weissman J.S., Hughes T.R., Strahl B.D., Grunstein M., Greenblatt J.F., Buratowski S., Krogan N.J. (2005) Cotranscriptional *Set2* methylation of histone H3 lysine 36 recruits a repressive *Rpd3* complex. *Cell*. 123, 593-605.

Khatib H., Zaitoun I., and Kim E.S. (2007). Comparative analysis of sequence characteristics of imprinted genes in human, mouse, and cattle. *Mamm. Genome* 18, 538-47.

Kielman M.F., Rindapää M., Gaspar C., van Poppel N., Breukel C., van Leeuwen S., Taketo M.M., Roberts S., Smits R., Fodde R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nat. Genet*. 32, 594-605.

Kim J. and Kim J.D. (2008). In vivo YY1 knockdown effects on genomic imprinting. *Hum. Mol. Genet.* 17, 391-401.

Kim J.D., Hinz A.K., Bergmann A., Huang J.M., Ovcharenko I., Stubbs L., and Kim J. (2006). Identification of clustered YY1 binding sites in imprinting control regions. *Genome Res.* 16, 901-911.

Kim T.H., Abdullaev Z.K., Smith A.D., Ching K.A., Loukinov D.I., Green R.D., Zhang M.Q., Lobanenkov V.V., Ren B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128: 1231–1245.

Knoepfler P.S. and Eisenman R.N. (1999). *Sin* meets NuRD and other tails of repression. *Cell* 99, 447-450.

Kohlmaier A., Savarese F., Lachner M., Martens J., Jenuwein T., and Wutz A. (2004). A chromosomal memory triggered by *Xist* regulates histone methylation in X inactivation. *PLoS Biol.* 2, E171.

Kouzarides T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

Kuramochi-Miyagawa S., Kimura T., Yomogida K., Kuroiwa A., Tadokoro Y., Fujita Y., Sato M., Matsuda Y., Nakano T. (2001). Two mouse *piwi*-related genes: *miwi* and *mili*. *Mech Dev*. 108, 121-33.

Kuroiwa Y., Kaneko-Ishino T., Kagitani F., Kohda T., Li L.L., Tada M., Suzuki R., Yokoyama M., Shiroishi T., Wakana S., Barton S.C., Ishino F.*, Surani M.A. (1996) Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. Nat. Genet.* 12:186-90.

Kurukuti S., Tiwari V.K., Tavoosidana G., Pugacheva E., Murrell A., Zhao Z., Lobanenkov V., Reik W., Ohlsson R. (2006). CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc Natl Acad Sci U S A* 103, 10684-9.

La Salle S., Mertineit C., Taketo T., Moens P.B., Bestor T.H., and Trasler J.M. (2004). Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev. Biol.* 268, 403-415.

Laird P.W., Jackson-Grusby L., Fazeil A., Dickinson S.L., Jung W.E., Li E., Weinberg R.A., and Jaenisch R. (1995). Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* 81, 197-205.

Lalande M. (1996). Parental imprinting and human disease. *Annu. Rev. Genet*.

30, 173-95.

Lee J.T., Davidow L.S., Warshawsky D. (1999). *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nat. Genet*. 21, 400-4.

Lees-Murdock D.J., De Felici M., and Walsh C.P. (2003). Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* 82, 230-237.

Lees-Murdock D.J., Shovlin T.C., Gardiner T., De Felici M., and Walsh C.P. (2005). DNA methyltransferase expression in the mouse germ line during periods of de novo methylation. *Dev. Dyn.* 232, 992-1002.

Lehnertz B., Ueda Y., Derijck A.A., Braunschweig U., Perez-Burgos L., Kubicek S., Chen T., Li E., Jenuwein T., and Peters A.H. (2003). *Suv39h*-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* 13, 1192-1200.

Lewis A., Mitsuya K., Umlauf D., Smith P., Dean W., Walter J., Higgins M., Feil R., and Reik W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat. Genet.* 36, 1291-1295.

Li B., Carey M., and Workman J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707-719.

Li E., Beard C., and Jaenisch R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365.

Lillycrop K.A., Phillips E.S., Jackson A.A., Hanson M.A., Burdge G.C. (2005). Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J. Nutr.* 135, 1382-6.

Lillycrop K.A., Slater-Jefferies J.L., Hanson M.A., Godfrey K.M., Jackson A.A., and Burdge G.C. (2007). Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. *Br. J. Nutr.* 97, 1064-1073.

Lillycrop K.A., Phillips E.S., Torrens C., Hanson M.A., Jackson A.A., and Burdge G.C. (2008). Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPARalpha promoter of the offspring. *Br. J. Nutr.*, 1-5.

Lindroth A.M., Park Y.J., McLean C.M., Dokshin G.A., Persson J.M., Herman H., Pasini D., Miró X., Donohoe M.E., Lee J.T., Helin K., Soloway P.D. (2008) Antagonism between DNA and H3K27 methylation at the imprinted *Rasgrf1* locus. *PLoS Genet*. 4, e1000145.

Ling J.Q., Li T., Hu J.F., Vu T.H., Chen H.L., Qiu X.W., Cherry A.M., and

Hoffman A.R. (2006). CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science* 312, 269-272.

Linglart A., Gensure R.C., Olney R.C., Juppner H., Bastepe M. (2005). A novel STX16 deletion in autosomal dominant pseudohypoparathyroidism type Ib redefines the boundaries of a cis-acting imprinting control element of GNAS. *Am. J. Hum. Genet.* 76: 804–814.

Lippman Z., Gendrel A.V., Black M., Vaughn M.W., Dedhia N., McCombie W.R., Lavine K., Mittal V., May B., Kasschau K.D., Carrington J.C., Doerge R.W., Colot V., Martienssen R. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471-476.

Loukinov D.I., Pugacheva E., Vatolin S., Pack S.D., Moon H., Chernukhin I., Mannan P., Larsson E., Kanduri C., Vostrov A.A., Cui H., Niemitz E.L., Rasko J.E., Docquier F.M., Kistler M., Breen J.J., Zhuang Z., Quitschke W.W., Renkawitz R., Klenova E.M., Feinberg A.P., Ohlsson R., Morse H.C., Lobanenkov V.V. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci. U S A* 99, 6806-6811.

Luedi P.P., Hartemink A.J., and Jirtle R.L. (2005). Genome-wide prediction of imprinted murine genes. *Genome Res.* 15, 875-884.

Luedi P.P., Dietrich F.S., Weidman J.R., Bosko J.M., Jirtle R.L., Hartemink

A.J. Computational and experimental identification of novel human imprinted genes. *Genome Res*. 17, 1723-30 (2007).

Lutz M., Burke L.J., Barreto G. Goeman F., Greb H., Arnold R., Schultheiss H., Brehm A., Kouzarides T., Lobanenkov V., Renkawitz R. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* 28, 1707-1713.

Ma D.K., Guo J.U., Ming G.L., Song H. (2009). DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle* 8, 1526-31.

Mackay D.J., Callaway J.L., Marks S.M., White H.E., Acerini C.L., Boonen S.E., Dayanikli P., Firth H.V., Goodship J.A., Haemers A.P., Hahnemann J.M., Kordonouri O., Masoud A.F., Oestergaard E., Storr J., Ellard S., Hattersley A.T., Robinson D.O., Temple I.K. (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat. Genet*. 40, 949-51.

Maeda N. and Hayashizaki Y. (2006). Genome-wide survey of imprinted genes. *Cytogenet. Genome Res*. 113, 144-52.

Mager J., Montgomery N.D., de Villena F.P., and Magnuson T. (2003). Genome imprinting regulated by the mouse Polycomb group protein *Eed*. *Nat. Genet.* 33, 502-507.

Malagnac F., Bartee L., and Bender J. (2002). An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* 21, 6842-6852.

Mancini-Dinardo D., Steele S.J., Levorse J.M., Ingram R.S., and Tilghman S.M. (2006). Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* 20, 1268-1282.

Mann J.R. and Lovell-Badge R.H. (1984). Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. *Nature*. 310, 66-7.

Martens J.H., O'Sullivan R.J., Braunschweig U., Opravil S., Radolf M., Steinlein P., and Jenuwein T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 24, 800-812.

Martienssen R.A. (2003). Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.* 35, 213-214.

Mathieu O., Probst A.V., and Paszkowski J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. *EMBO J.* 24, 2783-2791.

McGrath J. and Solter D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37, 179-183.

McLean C.M., Eilertson K.E., Bustamante C.D., Soloway P.D. (2009)

Successful computational prediction of novel imprinted genes. Submitted.

McLean C.M. and Soloway P.D. Microarray validation of candidate imprinted genes identified by computational prediction. Submitted.

McMahon A.P. and Bradley A. (1990). The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain. *Cell* 62, 1073-1085.

Mikkelsen T.S., Ku M., Jaffe D.B., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T.K., Koche R.P., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E.S., Bernstein B.E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.

Mizuno Y, Sotomaru Y, Katsuzawa Y, Kono T, Meguro M, Oshimura M, Kawai J, Tomaru Y, Kiyosawa H, Nikaido I, Amanuma H, Hayashizaki Y, Okazaki Y. (2002). Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem. Biophys. Res. Commun*. 290, 1499-505.

Morgan H.D., Santos F., Green K., Dean W., and Reik W. (2005). Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* 14, R47-R58.

Murrell A., Heeson S., and Reik W. (2004). Interaction between differentially

methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat. Genet.* 36, 889-893.

Nagano T., Mitchell J.A., Sanz L.A., Pauler F.M., Ferguson-Smith A.C., Feil R., Fraser P. (2008). The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717-20.

Ng K., Pullirsch D., Leeb M., and Wutz A. (2007). *Xist* and the order of silencing. *EMBO Reports* 8, 34-39.

Neumann B, Kubicka P, Barlow DP. (1995). Characteristics of imprinted genes. *Nat. Genet.* 9, 12-3.

Nikaido I., Saito C., Mizuno Y., Meguro M., Bono H., Kadomura M., Kono T., Morris G.A., Lyons P.A., Oshimura M., Hayashizaki Y., Okazaki Y.; RIKEN GER Group; GSL Members. (2003). Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res.* 13, 1402-9.

Ohlsson R., Nyström A,. Pfeifer-Ohlsson S., Töhönen V., Hedborg F., Schofield P., Flam F., Ekström T.J. (1993). IGF2 is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. *Nat Genet.* 4, 94-7.

Ooi S.K., Qiu C., Bernstein E., Li K., Jia D., Yang Z., Erdjument-Bromage H., Tempst P., Lin S.P., Allis C.D., Cheng X., Bestor T.H. (2007). DNMT3L

connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.

Oswald J., Engemann S., Lane N., Mayer W., Olek A., Fundele R., Dean W., Reik W., Walter J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol*. 20, 475-8.

Pant V., Mariano P., Kanduri C., Mattsson A., Lobanenkov V., Heuchel R., Ohlsson R. (2003) The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.* 17: 586–590.

Pant V., Kurukuti S., Pugacheva E., Shamsuddin S., Mariano P., Renkawitz R., Klenova E., Lobanenkov V., and Ohlsson R. (2004). Mutation of a single CTCF target site within the *H19* imprinting control region leads to loss of *Igf2* imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell Biol.* 24, 3497-3504.

Papp B. and Muller J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20, 2041-2054.

Pasini D., Bracken A.P., Jensen M.R., Lazzerini Denchi E., and Helin K. (2004). *Suz12* is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* 23, 4061-4071.

Paulsen M. and Ferguson-Smith A.C. (2001). DNA methylation in genomic imprinting, development, and disease. *J. Pathol.* 195, 97-110.

Pearsall R.S., Plass C., Romano M.A., Garrick M.D., Shibata H., Hayashizaki Y., Held W.A. (1999). A direct repeat sequence at the *Rasgrf1* locus and imprinted expression. *Genomics* 55, 194-201.

Perk J., Makedonski K., Lande L., Cedar H., Razin A., Shemer R. (2002) The imprinting mechanism of the Prader-Willi/Angelman regional control center. *EMBO J.* 21: 5807–5814.

Peters A.H., O'Carroll D., Scherthan H., Mechtler K., Sauer S., Schöfer C., Weipoltshammer K., Pagani M., Lachner M., Kohlmaier A., Opravil S., Doyle M., Sibilia M., Jenuwein T. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323–337.

Peters A.H., Kubicek S., Mechtler K., O'Sullivan R.J., Derijck A.A., Perez-Burgos L., Kohlmaier A., Opravil S., Tachibana M., Shinkai Y., Martens J.H., Jenuwein T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* 12, 1577-1589.

Plass C., Shibata H., Kalcheva I., Mullins L., Kotelevtseva N., Mullins J., Kato R., Sasaki H., Hirotsune S., Okazaki Y., Held W.A., Hayashizaki Y., Chapman V.M. (1996) Identification of *Grf1* on mouse chromosome 9 as an imprinted

gene by RLGS-M. *Nat. Genet.* 14, 106-9.

Plass C. and Soloway P.D. (2002) DNA methylation, imprinting, and cancer, *Eur. J. Hum. Genet.* 10, 6-16.

Pollard K.S., Serre D., Wang X., Tao H., Grundberg E., Hudson T.J., Clark A.G., Frazer K. (2008). A genome-wide approach to identifying novel-imprinted genes. *Hum. Genet.* 122, 625-34.

Ptashne M. (2007). On the use of the word 'epigenetic'. *Curr. Biol*. 17, R233-6.

Qiu X., Vu T.H., Lu Q., Ling J.Q., Li T., Hou A., Wang S.K., Chen H.L., Hu J.F., Hoffman A.R. (2008). A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal *Igf2* allele. *Mol. Endocrinol*. 22, 1476-88.

Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201-12.

Rand E., Ben-Porath I., Keshet I., and Cedar H. (2004). CTCF elements direct allele-specific undermethylation at the imprinted *H19* locus. *Curr. Biol.* 14, 1007-1012.

Rassoulzadegan M., Grandjean V., Gounon P., Vincent S., Gillot I., and Cuzin F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change

in the mouse. *Nature* 441, 469-474.

Reik W., Murrell A., Lewis A., Mitsuya K., Umlauf D., Dean W., Higgins M., and Feil R. (2004). Chromosome loops, insulators, and histone methylation: new insights into regulation of imprinting in clusters. *Cold Spring Harbor Symposia on Quantitative Biology* 69, 29-37.

Reinhart B., Eljanne M., and Chaillet J.R. (2002). Shared role for differentially methylated domains of imprinted genes. *Mol. Cellular Biol.* 22, 2089- 2098.

Ro S., Park C., Song R., Nguyen D., Jin J., Sanders K.M., McCarrey J.R., Yan W. (2007). Cloning and expression profiling of testis-expressed piRNA-like RNAs. *RNA*. 13, 1693-702.

Robertson K.D. (2005). DNA methylation and human disease. *Nat. Reviews* 6, 597- 610.

Rowitch D.H., Echelard Y., Danielian P.S., Gellner K., Brenner S., McMahon A.P. (1998) Identification of an evolutionarily conserved 110 base-pair *cis*-acting regulatory sequence that governs *Wnt-1* expression in the murine neural plate. *Development* 125, 2735-46.

Ruf N., Bähring S., Galetzka D., Pliushch G., Luft F.C., Nürnberg P., Haaf T., Kelsey G., Zechner U. (2007). Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Hum. Mol. Genet.* 16, 2591-9.

Santos, F. and Dean W. (2004). Epigenetic reprogramming during early development in mammals. *Reproduction* 127, 643-651.

Sarg B., Koutzamani E., Helliger W., Rundquist I., and Lindner H.H. (2002). Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J. Biol. Chem.* 277, 39195-39201.

Satijn D.P., Hamer K.M., den Blaauwen J., and Otte A.P. (2001). The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Mol. Cellular Biol.* 21, 1360-1369.

Schlesinger Y., Straussman R., Keshet I., Farkash S., Hecht M., Zimmerman J., Eden E., Yakhini Z., Ben-Shushan E., Reubinoff B.E., Bergman Y., Simon I., Cedar H. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* 39, 232-236.

Schoenherr C.J., Levorse J.M., and Tilghman S.M. (2003). CTCF maintains differential methylation at the *Igf2/H19* locus. *Nat. Genet.* 33, 66-69.

Schulz R., Menheniott T.R., Woodfine K., Wood A.J., Choi J.D., Oakey R.J. (2006). Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucl. Acids Res*. 34, e88.

Seitz H., Royo H., Bortolin M.L., Lin S.P., Ferguson-Smith A.C., Cavaillé J. (2004). A large imprinted microRNA gene cluster at the mouse *Dlk1-Gtl2*

domain. *Genome Res*. 14, 1741-8.

Shao W.J., Tao L.Y., Gao C., Xie J.Y., Zhao R.Q. (2008). Alterations in methylation and expression levels of imprinted genes *H19* and *Igf2* in the fetuses of diabetic mice. *Comp. Med*. 58, 341-6.

Shemer R., Birger Y., Dean W.L., Reik W., Riggs A.D., and Razin A. (1996). Dynamic methylation adjustment and counting as part of imprinting mechanisms. *Proc. Natl. Acad. Sci. U S A* 93, 6371-6376.

Shibata H., Yoda Y., Kato R., Ueda T., Kamiya M., Hiraiwa N., Yoshiki A., Plass C., Pearsall R.S., Held W.A., Muramatsu M., Sasaki H., Kusakabe M., and Hayashizaki Y. (1998). A methylation imprint mark in the mouse imprinted gene *Grf1/Cdc25Mm* locus shares a common feature with the *U2afbp-rs* gene: An association with a short tandem repeat and a hypermethylated region. *Genomics* 49, 30-37.

Shin J.Y., Fitzpatrick G.V., Higgins M.J. (2008). Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *EMBO J*. 2008 27, 168-78.

Shumaker D.K., Dechat T., Kohlmaier A., Adam S.A., Bozovsky M.R., Erdos M.R., Eriksson M., Goldman A.E., Khuon S., Collins F.S., Jenuwein T., Goldman R.D. (2006). Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc. Natl. Acad. Sci. U S A* 103, 8703-8708.

Sinkkonen L., Hugenschmidt T., Berninger P., Gaidatzis D., Mohn F., Artus-Revel C.G., Zavolan M., Svoboda P., Filipowicz W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol*. 15, 259-67.

Sleutels F., Zwart R., and Barlow D.P. (2002). The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810-813.

Smallwood A., Esteve P.O., Pradhan S., and Carey M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & Development* 21, 1169-1178.

Smith R.J., Dean W., Konfortova G., Kelsey G. (2003). Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res*. 13:558-69.

Smith S.S., Laayoun A., Lingeman R.G., Baker D.J., Riley J. (1994). Hypermethylation of telomere-like foldbacks at codon 12 of the human c-Ha-ras gene and the trinucleotide repeat of the FMR-1 gene of fragile X. *J. Mol. Biol*. 243, 143-51.

Soppe W.J., Jasencakova Z., Houben A., Kakutani T., Meister A., Huang M.S., Jacobsen S.E., Schubert I., and Fransz P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* 21, 6549-6559.

Splinter E., Heath H., Kooren J., Palstra R.J., Klous P., Grosveld F., Galjart N., de Laat W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 20, 2349-54.

Srinivasan L., Pan X., and Atchison M.L. (2005). Transient requirements of YY1 expression for PcG transcriptional repression and phenotypic rescue. *J. Cell. Biochem.* 96, 689-699.

Srivastava M., Hsieh S., Grinberg A., Williams-Simons L., Huang S.P., and Pfeifer K. (2000). *H19* and *Igf2* monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19. *Genes Dev.* 14, 1186-1195.

Srivastava M., Frolova E., Rottinghaus B., Boe S.P., Grinberg A., Lee E., Love P.E., Pfeifer K. (2003) Imprint control element-mediated secondary methylation imprints at the *Igf2/H19* locus. *J. Biol. Chem.* 278: 5977–5983.

Stöger R., Kubicka P., Liu C.G., Kafri T., Razin A., Cedar H., Barlow D.P. (2003). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* 73, 61-71.

Surani M.A., Barton S.C., and Norris M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308, 548-550.

Swain J.L., Stewart T.A., Leder P. (1987). Parental legacy determines

methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. *Cell* 50, 719-27.

Tamaru H. and Selker E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* 414, 277-283.

Tamaru H., Zhang X., McMillen D., Singh P.B., Nakayama J., Grewal S.I., Allis C.D., Cheng X., and Selker E.U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nat. Genet.* 34, 75-79.

Tanimoto K., Shimotsuma M., Matsuzaki H., Omori A., Bungert J., Engel J.D., Fukamizu A. (2005). Genomic imprinting recapitulated in the human beta-globin locus. *Proc. Natl. Acad. Sci. U S A* 102, 10250-10255.

Tariq M., Saze H., Probst A.V., Lichota J., Habu Y., and Paszkowski J. (2003). Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc. Natl. Acad. Sci. U S A* 100, 8823-8827.

Thakur N., Tiwari V.K., Thomassin H., Pandey R.R., Kanduri M., Gondor A., Grange T., Ohlsson R., and Kanduri C. (2004). An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol. Cell. Biol.* 24, 7855-7862.

Thorvaldsen J.L., Duran K.L., and Bartolomei M.S. (1998). Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* 12, 3693-3702.

Toyota M., Ahuja N., Ohe-Toyota M., Herman J.G., Baylin S.B., and Issa J.P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U S A* 96, 8681-6.

Vagin V.V., Sigova A., Li C., Seitz H., Gvozdev V., Zamore P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320-4.

Verdel A., Jia S., Gerber S., Sugiyama T., Gygi S., Grewal S.I., and Moazed D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303, 672-676.

Vire E., Brenner C., Deplus R., Blanchon L., Fraga M., Didelot C., Morey L., Van Eynde A., Bernard D., Vanderwinden J.M., Bollen M., Esteller M., Di Croce L., de Launoit Y., Fuks F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871-874.

Waddington C.H. (1942). The epigenotype, Vol 1 (London).

Wagschal A., Sutherland H.G., Woodfine K., Henckel A., Chebli K., Schulz R., Oakey R.J., Bickmore W.A., and Feil R. (2008). G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol. and Cell. Biol.* 28, 1104-1113.

Walter J., Hutter B., Khare T., and Paulsen M. (2006). Repetitive elements in

imprinted genes. *Cytogenet. Genome Res.* 113, 109-115.

Wang X., Sun Q., McGrath S.D., Mardis E.R., Soloway P.D., Clark A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS ONE*. 3, e3839.

Wassenegger M., Heimes S., Riedel L., Sänger H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. *Cell* 76, 567-76.

Waterland R.A. and Jirtle R.L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* 23, 5293-5300.

Wen B., Wu H., Bjornsson H., Green R.D., Irizarry R., Feinberg A.P. (2008). Overlapping euchromatin/heterochromatin- associated marks are enriched in imprinted gene regions and predict allele-specific modification. *Genome Res*. 18, 1806-13.

Wilkinson D.G., Bailes J.A., and McMahon A.P. (1987). Expression of the proto-oncogene *int-1* is restricted to specific neural cells in the developing mouse embryo. *Cell* 50, 79-88.

Wilkinson F.H., Park K., and Atchison M.L. (2006). Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc. Natl. Acad. Sci. U S A* 103, 19296-19301.

Wolf J.B., Cheverud J.M., Roseman C., and Hager R. (2008). Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet.* 4, e1000091.

Wu M.Y., Tsai T.F., Beaudet A.L. (2006). Deficiency of *Rbbp1/Arid4a* and *Rbbp1l1/Arid4b* alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. *Genes Dev*. 20, 2859-70.

Wutz A., Smrzka O.W., Schweifer N., Schellander K., Wagner E.F., and Barlow D.P. (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* 389, 745-749.

Wutz A., Rasmussen T.P., and Jaenisch R. (2002). Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* 30, 167-174.

Xie T., Chen M., Gavrilova O., Lai E.W., Liu J., Weinstein L.S. (2008). Severe obesity and insulin resistance due to deletion of the maternal *Gsalpha* allele is reversed by paternal deletion of the *Gsalpha* imprint control region. *Endocrinology*. 149, 2443-50.

Xin Z., Tachibana M., Guggiari M., Heard E., Shinkai Y., and Wagstaff J. (2003). Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. *J. Biol. Chem.* 278, 14996- 15000.

Xiong Z. and Laird P.W. (1997). COBRA: a sensitive and quantitative DNA

methylation assay. *Nucl. Acids Res.* 25, 2532-2534.

Yamazaki Y., Mann M.R.W., Lee S.S., Marh J., McCarrey J.R., Yanagimachi R., and Bartolomei M.S. (2003). Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. *Proc. Natl. Acad. Sci. U S A* 100, 12207-12212.

Yoon B.J., Herman H., Sikora A., Smith L.T., Plass C., and Soloway P.D. (2002). Regulation of DNA methylation of *Rasgrf1*. *Nat. Genet.* 30, 92-96.

Yoon B., Herman H., Hu B., Park Y.J., Lindroth A., Bell A., West A.G., Chang Y., Stablewski A., Piel J.C., Loukinov D.I., Lobanenkov V.V., Soloway P.D. (2005). *Rasgrf1* imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. *Mol. Cell. Biol.* 25, 11184-11190.

Yusufzai T.M., Tagami H., Nakatani Y., Felsenfeld G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Moli Cell*. 13, 291-8.

Zaratiegui M., Irvine D.V., and Martienssen R.A. (2007). Noncoding RNAs and gene silencing. *Cell* 128, 763-776.

Zhang X., Clarenz O., Cokus S., Bernatavichute Y.V., Pellegrini M., Goodrich J., and Jacobsen S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* 5, e129.

Zhao J., Sun B.K., Erwin J.A., Song J.J., Lee J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750-6.